

1 Topology optimization by distribution of isotropic material

In this chapter we present an overview of the basic ingredients of what we will denote as the *material distribution method* for finding the optimum lay-out of a linearly elastic structure. In this context the “lay-out” of the structure includes information on the topology, shape and sizing of the structure and the material distribution method allows for addressing all three problems simultaneously.

Sizing, shape, and topology optimization problems address different aspects of the structural design problem. In a typical *sizing* problem the goal may be to find the optimal thickness distribution of a linearly elastic plate or the optimal member areas in a truss structure. The optimal thickness distribution minimizes (or maximizes) a physical quantity such as the mean compliance (external work), peak stress, deflection, etc., while equilibrium and other constraints on the state and design variables are satisfied. The design variable is the thickness of the plate and the state variable may be its deflection. The main feature of the sizing problem is that the domain of the design model and state variables is known *a priori* and is fixed throughout the optimization process. On the other hand, in a *shape* optimization problem the goal is to find the optimum shape of this domain, that is, the shape problem is defined on a domain which is now the design variable. *Topology* optimization of solid structures involves the determination of features such as the number and location and shape of holes and the connectivity of the domain.

1.1 Problem formulation and parametrization of design

The lay-out problem that shall be defined in the following combines several features of the traditional problems in structural design optimization. The purpose of topology optimization is to find the optimal lay-out of a structure within a specified region. The only known quantities in the problem are the applied loads, the possible support conditions, the volume of the structure to be constructed and possibly some additional design restrictions such as the location and size of prescribed holes or solid areas. In this problem the physical size and the shape and connectivity of the structure are unknown.

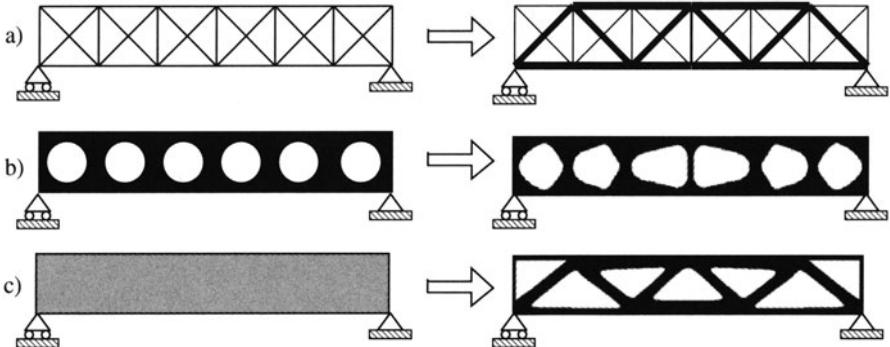


Fig. 1.1. Three categories of structural optimization. a) Sizing optimization of a truss structure, b) shape optimization and d) topology optimization. The initial problems are shown at the left hand side and the optimal solutions are shown at the right

The topology, shape, and size of the structure are not represented by standard parametric functions but by a set of distributed functions defined on a *fixed design domain*. These functions in turn represent a parametrization of the stiffness tensor of the continuum and it is the suitable choice of this parametrization which leads to the proper design formulation for topology optimization.

1.1.1 Minimum compliance design

In the following, the general set-up for optimal shape design formulated as a *material distribution problem* is described. The set-up is analogous to well known formulations for sizing problems for discrete and continuum structures [1], and to truss topology design formulations that are described later in this monograph. It is important to note that the problem type we will consider is from a computational point of view inherently large scale, both in state and in the design variables. For this reason the first problems treated in this area employed the simplest type of design problem formulation in terms of objective and constraint, namely designing for minimum compliance (maximum global stiffness) under simple resource constraints. This is also conceptually a natural starting point for this exposition as its solution reflects many of the fundamental issues in the field.

Consider a mechanical element as a body occupying a domain Ω^{mat} which is part of a larger reference domain Ω in \mathbf{R}^2 or \mathbf{R}^3 . The reference domain Ω is chosen so as to allow for a definition of the applied loads and boundary conditions and the reference domain is sometimes called the ground structure, in parallel with terminology used in truss topology design. Referring to the reference domain Ω we can define the optimal design problem as the problem

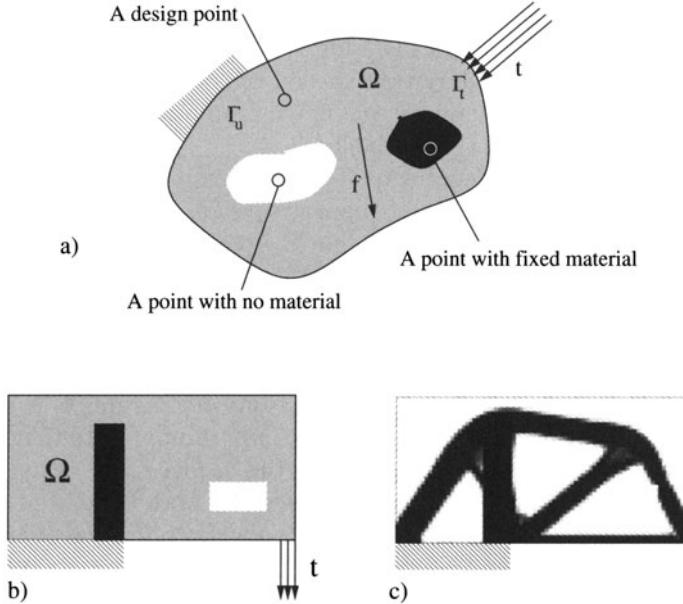


Fig. 1.2. a) The generalized shape design problem of finding the optimal material distribution in a two-dimensional domain. b) Example rectangular design domain and c) topology optimized solution based on a 3200 element discretization and 50% material volume.

of finding the optimal choice of stiffness tensor $E_{ijkl}(x)$ ¹ which is a variable over the domain. Introducing the energy bilinear form (i.e., the internal virtual work of an elastic body at the equilibrium u and for an arbitrary virtual displacement v):

$$a(u, v) = \int_{\Omega} E_{ijkl}(x) \varepsilon_{ij}(u) \varepsilon_{kl}(v) d\Omega ,$$

with linearized strains $\varepsilon_{ij}(u) = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$ and the load linear form

$$l(u) = \int_{\Omega} f u d\Omega + \int_{\Gamma_T} t u ds ,$$

the minimum compliance (maximum global stiffness) problem takes the form

$$\begin{aligned} & \min_{u \in U, E} l(u) \\ & \text{s.t. : } a_E(u, v) = l(v), \quad \text{for all } v \in U , \\ & \quad E \in \mathbf{E}_{ad} . \end{aligned} \tag{1.1}$$

¹ In what follows we use a standard tensor notation consistent with a Cartesian reference frame; this does not imply a loss of generality.

Here the equilibrium equation is written in its weak, variational form, with U denoting the space of kinematically admissible displacement fields, f are the body forces and t the boundary tractions on the traction part $\Gamma_T \subset \Gamma \equiv \partial\Omega$ of the boundary. Note that we use the index E to indicate that the bilinear form a_E depends on the design variables.

In problem (1.1), \mathbf{E}_{ad} denotes the set of admissible stiffness tensors for our design problem. In the case of topology design, \mathbf{E}_{ad} could, for example, consist of all stiffness tensors that attain the material properties of a given isotropic material in the (unknown) set Ω^{mat} and zero properties elsewhere, the limit of resource being expressed as $\int_{\Omega^{mat}} 1 d\Omega \leq V$. The various possible definitions of \mathbf{E}_{ad} is the subject of the following section.

When solving problems of the form (1.1) by computational means a typical approach, and the one used throughout this monograph, is to discretize the problem using finite elements. It is here important to note that there are *two* fields of interest in (1.1), namely both the displacement u and the stiffness E . If we use the same finite element mesh for both fields, and discretize E as constant in each element, we can write the discrete form of (1.1) as

$$\begin{aligned} & \min_{\mathbf{u}, E_e} \mathbf{f}^T \mathbf{u} \\ & \text{s.t. : } \mathbf{K}(E_e) \mathbf{u} = \mathbf{f} , \\ & \quad E_e \in \mathbf{E}_{ad} . \end{aligned} \tag{1.2}$$

Here \mathbf{u} and \mathbf{f} are the displacement and load vectors, respectively. The stiffness matrix \mathbf{K} depends on the stiffness E_e in element e , numbered as $e = 1, \dots, N$, and we can write \mathbf{K} in the form

$$\mathbf{K} = \sum_{e=1}^N \mathbf{K}_e(E_e) ,$$

where \mathbf{K}_e is the (global level) element stiffness matrix.

1.1.2 Design parametrization

In the design of the topology of a structure we are interested in the determination of the optimal placement of a given isotropic material in space, i.e., we should determine which points of space should be material points and which points should remain void (no material). That is, we think of the geometric representation of a structure as similar to a black-white rendering of an image. In discrete form this then corresponds to a black-white raster representation of the geometry, with “pixels” (or “voxels”) given by the finite element discretization.

Restricting our spatial extension to the reference domain Ω , we are thus seeking to determine the optimal subset Ω^{mat} of material points. For the

optimization problem defined above, this approach implies that the set E_{ad} of admissible stiffness tensors consists of those tensors for which²:

$$E_{ijkl} = 1_{\Omega^{\text{mat}}} E_{ijkl}^0, \quad 1_{\Omega^{\text{mat}}} = \begin{cases} 1 & \text{if } x \in \Omega^{\text{mat}}, \\ 0 & \text{if } x \in \Omega \setminus \Omega^{\text{mat}}, \end{cases}, \quad (1.3)$$

$$\int_{\Omega} 1_{\Omega^{\text{mat}}} d\Omega = \text{Vol}(\Omega^{\text{mat}}) \leq V.$$

Here the last inequality expresses a limit, V , on the amount of material at our disposal, so that the minimum compliance design is for a limited (fixed) volume. The tensor E_{ijkl}^0 is the stiffness tensor for the given *isotropic* material and one normally writes $E_{ijkl} \in L^\infty(\Omega)$ to indicate the relevant function-space for our problem. Note that this definition of E_{ad} means that we have formulated a distributed, discrete valued design problem (a 0-1 problem).

The most commonly used approach to solve this problem is to replace the integer variables with continuous variables and then introduce some form of penalty that steers the solution to discrete 0-1 values³. The design problem for the fixed domain is then formulated as a sizing problem by modifying the stiffness matrix so that it depends continuously on a function which is interpreted as a density of material [6]. This function is then the design variable. The requirement is that the optimization results in designs consisting almost entirely of regions of material or no material. This means that intermediate values of this artificial density function should be penalized in a manner analogous to other continuous optimization approximations of 0-1 problems.

One possibility which has proven very popular and extremely efficient is the so-called penalized, proportional stiffness model (the SIMP-model⁴):

$$E_{ijkl}(x) = \rho(x)^p E_{ijkl}^0, \quad p > 1,$$

$$\int_{\Omega} \rho(x) d\Omega \leq V; \quad 0 \leq \rho(x) \leq 1, \quad x \in \Omega, \quad (1.4)$$

Here the “density” $\rho(x)$ is the design function and E_{ijkl}^0 represents the material properties of a given isotropic material. One refers to ρ as a density of material by the fact that the volume of the structure is evaluated as $\int_{\Omega} \rho(x) d\Omega$. The density interpolates between the material properties 0 and E_{ijkl}^0 :

$$E_{ijkl}(\rho = 0) = 0, \quad E_{ijkl}(\rho = 1) = E_{ijkl}^0,$$

² We consider isotropic materials only, as for anisotropic materials the placement of the principal directions of the material should also be considered as a design variable.

³ Methods that deals directly with the integer problem are briefly discussed in Sect. 1.5.6.

⁴ SIMP: Solid Isotropic Material with Penalization.

meaning that if a final design has density zero or one in all points, this design is a black-and-white design for which the performance has been evaluated with a correct physical model. In SIMP one will choose to use $p > 1$ so that intermediate densities are unfavourable in the sense that the stiffness obtained is small compared to the cost (volume) of the material. In other words, by specifying a value of p higher than one makes it “uneconomical” to have intermediate densities in the optimal design. Thus the penalization is achieved without the use of any explicit penalization scheme. For problems where the volume constraint is active, experience shows that optimization does actually result in such designs if one chooses p sufficiently big (in order to obtain true “0-1” designs, $p \geq 3$ is usually required). Moreover, it has been proven for the minimum compliance problem in *discrete* form (cf., problem (1.2)) that for p large enough there exists a globally optimal solution of 0-1 form, provided the volume constraint is compatible with such a design (Rietz 2001) (see also section 1.5.4). The SIMP interpolation is the basis for most computational results in the first half of this monograph.

We note that the original “0-1” problem is defined on a fixed reference domain and this together with the SIMP-interpolation means that the optimal topology problem takes on the form of a standard sizing problem on a fixed domain.

It has often been questioned if the SIMP-model can be interpreted in physical terms (the term “material” is part of the acronym!). That is, can one find a material, for example as a composite, which realizes the interpolation model. It is important to point out that this comparison of an interpolation scheme like SIMP with micromechanical models is significant mainly for the benefit of understanding the nature of such computational measures. If a numerical scheme leads to black-and-white designs one can in essence choose to ignore the physical relevance of intermediate steps which may include “grey”. However, the question of physical relevance is often raised, especially as most computational schemes involving interpolations do give rise to designs which are not *completely* clear of “grey”. Also, the physical realization of all feasible designs plays a role when interpreting results from a premature termination of an optimization algorithm that has not converged fully to a 0-1 design.

We will return later in section 2.10 to the construction of a material model that mimics the SIMP interpolation model. Central in such considerations is a comparison with the Hashin-Shtrikman bounds for two-phase materials, which expresses the limits of possible *isotropic* material properties one can achieve by constructing composites (materials with microstructure) from two given, linearly elastic, isotropic materials [4]. Without further elaboration (and referring the reader to section 1.5.4) we remark here that the SIMP-model can indeed be considered as a material model if the power p satisfies that:

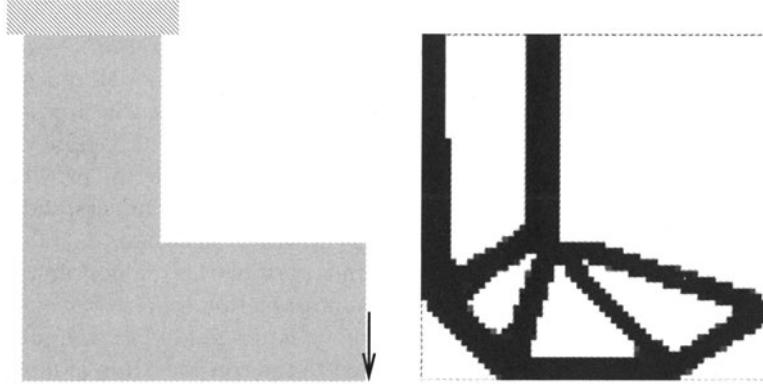


Fig. 1.3. A black-and-white minimum compliance design for a loaded knee structure obtained with the SIMP interpolation scheme. The discretization is 60 by 60 elements and the material volume is limited to 47% of the design domain.

$$\begin{aligned} p &\geq \max \left\{ \frac{2}{1-\nu^0}, \frac{4}{1+\nu^0} \right\} \quad (\text{in 2-D}), \\ p &\geq \max \left\{ 15 \frac{1-\nu^0}{7-5\nu^0}, \frac{3}{2} \frac{1-\nu^0}{1-2\nu^0} \right\} \quad (\text{in 3-D}), \end{aligned} \quad (1.5)$$

where ν^0 is the Poisson ratio of the given base material with stiffness tensor E_{ijkl}^0 (Bendsøe & Sigmund 1999). In dimension 2 this implies that the smallest p is 3, which is admissible for $\nu^0 = 1/3$. In dimension 3 the smallest admissible p is 2, but for $\nu^0 = 1/3$ one should also in 3-D choose p greater than 3.

The use of the SIMP interpolation scheme addresses the integer format of the original setting of the topology design problem with designs given as in (1.3). It converts this integer problem to a sizing problem that typically results in what for all practical purposes can be considered as 0-1 designs. Another serious problem associated with the 0-1 problem statement, and a problem SIMP does not resolve, is the now well established lack of existence of solutions to the distributed problem [3], [25], [34]. This is not only a serious theoretical drawback. It also has the effect of making the computational results sensitive to the fineness of the finite element mesh discretization. As mentioned, the interpolation scheme does not directly resolve this problem, and further considerations are in place to assure a well-posed distributed design problem that also is benign in terms of mesh-refinement of the finite element model. This aspect will be covered in some detail in 1.3.1.

The role of composites The initial work on numerical methods for topology design of continuum structures was based on using composite materials as the basis for describing varying material properties in space [3], [6]. This approach has been named the *homogenization method* and its development was strongly inspired by theoretical studies on generalized shape design in

conduction and torsion problems and by numerical and theoretical work related to plate design (see, e.g., [25], [29]). The homogenization approach to topology design can also be viewed as an interpolation model and many of the developments in the area of topology design of structures are based on this type of interpolation schemes. However, composites in topology design has deeper roots than this and composites play a key role for providing insight in the optimal use of materials in the large and for understanding the mathematics and the physics of the “0-1” design problem.

We have here chosen to base the first part of this monograph on the use of the somewhat simpler SIMP-type interpolation schemes (see also section 1.5.4). This allows for a concentration on issues related to computational implementations and developments related to the consideration of more elaborate optimal design settings than the minimum compliance design problem. In the second half of the monograph we will return to discuss issues related to the optimal use of material in general and the details of the homogenization based method in particular.

1.1.3 Alternative problem forms

For the developments in the following it is important to note that problem (1.1) can be given a number of equivalent formulations. These reformulations use the special structure of the minimum compliance problem and they are extremely useful for analysis and the development of specialized computational procedures for this type of problem. For this purpose we note that the equilibrium condition of problem (1.1) can be expressed in terms of the principle of minimum potential energy. That is, the displacement field u is a minimizer of the functional $F(v) = \frac{1}{2}a_E(v, v) - l(v)$ on U (the total potential energy). Then note that the value $F(u)$ of the potential energy at equilibrium equals $-\frac{1}{2}l(u) < 0$. Thus problem (1.1) can be written as

$$\max_{E \in \mathbb{E}_{\text{ad}}} \min_{u \in U} \left\{ \frac{1}{2}a_E(u, u) - l(u) \right\}. \quad (1.6)$$

Problem (1.6) can also be formulated in terms of stresses. Expressing the inner equilibrium problem of (1.6) in terms of the (dual) principle of minimum complementary energy, we have the formulation

$$\min_{E \in \mathbb{E}_{\text{ad}}} \min_{\sigma \in S} \left\{ \frac{1}{2} \int_{\Omega} C_{ijkl} \sigma_{ij} \sigma_{kl} d\Omega \right\}, \quad (1.7)$$

of the minimum compliance design problem. Here $C_{ijkl} = (E^{-1})_{ijkl}$ is the compliance tensor, and the minimization with respect to the stresses σ is taken over the set S of statically admissible stress fields, i.e.

$$S = \{\sigma \mid \operatorname{div} \sigma + f = 0 \text{ in } \Omega, \sigma \cdot n = t \text{ on } \Gamma_T\}.$$

From the problem statement (1.6) we see that if the displacement field in an optimal structure is known, then the optimal distribution of the stiffness is such that the strain energy is maximized. Likewise, for a known stress distribution in the optimal structure the complementary energy is minimized, cf., (1.7). This characterization plays an important role in understanding the nature of the design problem, both theoretically and computationally.

1.2 Solution methods

The use of an interpolation scheme like SIMP allows us to convert the optimal topology problem into a *sizing* problem on a *fixed* domain. Compared to many traditional sizing problems for, e.g., frames and built-up structures of plates, stringers, etc., the present problem differs in that the number of design variables is typically very large (the number of design parameters and the number of analysis variables is of the same order of magnitude). Thus efficiency of the optimization procedure is crucial and one typically has to adopt optimization settings that trade number of constraints for number of design variables. The compliance design problem is an example of this. One can here work with many variables, as the optimization problem (seen as a problem in the density only) has just one constraint in addition to the simple box-constraints giving upper and lower limits on the density variable.

1.2.1 Conditions of optimality

In the following we shall derive the necessary conditions of optimality for the density ρ of the minimum compliance design problem that employs the SIMP interpolation scheme.

Following standard optimality criteria methods used in structural optimization [7], the simple structure of the continuum, single load problem (1.1) can be utilized to generate extremely efficient computational update schemes for solving the problems we address here. The key is to devise iterative methods which, for a previously computed design and its associated displacements, update the design variables at each point (or rather at each element of a finite element discretization) independently from the updates at other points, based on the necessary conditions of optimality. To this end, we first recapitulate the form of the minimum compliance problem (1.1) written for the case of the SIMP interpolation. In the continuum setting this is

$$\begin{aligned} & \min_{u \in U, \rho} l(u) \\ \text{s.t. : } & a_E(u, v) = l(v), \quad \text{for all } v \in U, \\ & E_{ijkl}(x) = \rho(x)^p E_{ijkl}^0, \\ & \int_{\Omega} \rho(x) d\Omega \leq V; \quad 0 < \rho_{\min} \leq \rho \leq 1. \end{aligned} \tag{1.8}$$

Note that we here have introduced a lower bound ρ_{\min} on the density in order to prevent any possible singularity of the equilibrium problem. In typical applications we set $\rho_{\min} = 10^{-3}$.

With Lagrange multipliers $\Lambda, \lambda^-(x), \lambda^+(x)$ for the constraints of (1.8), the necessary conditions for optimality for the sizing variable ρ are a subset of the stationarity conditions for the Lagrange function

$$\begin{aligned}\mathcal{L} = & l(u) - \left\{ a_E(u, \bar{u}) - l(\bar{u}) \right\} + \Lambda \left(\int_{\Omega} \rho(x) d\Omega - V \right) + \\ & \int_{\Omega} \lambda^+(x)(\rho(x) - 1) d\Omega + \int_{\Omega} \lambda^-(x)(\rho_{\min} - \rho(x)) d\Omega ,\end{aligned}$$

where \bar{u} is the Lagrange multiplier for the equilibrium constraint. Note that \bar{u} belongs to the set U of kinematically admissible displacement fields. Under the assumption that $\rho \geq \rho_{\min} > 0$ (so that displacement fields are unique), the conditions for optimality with respect to variations of the displacement field u give that $\bar{u} = u$ while the condition for ρ becomes:

$$\frac{\partial E_{ijkl}}{\partial \rho} \varepsilon_{ij}(u) \varepsilon_{kl}(u) = \Lambda + \lambda^+ - \lambda^- , \quad (1.9)$$

with the switching conditions

$$\lambda^- \geq 0, \lambda^+ \geq 0, \lambda^-(\rho_{\min} - \rho(x)) = 0, \lambda^+(\rho(x) - 1) = 0 . \quad (1.10)$$

For intermediate densities ($\rho_{\min} < \rho < 1$) the conditions (1.9), can be written as

$$p\rho(x)^{p-1} E_{ijkl}^0 \varepsilon_{ij}(u) \varepsilon_{kl}(u) = \Lambda , \quad (1.11)$$

which expresses that the strain energy density-like left-hand side term ⁵ is constant and equal to Λ for all intermediate densities. This is thus a condition that is similar to the fully stressed design condition in plastic design. As we expect areas with high energy to be too low on stiffness, we devise the following fix-point type update scheme for the density [7]:

$$\rho_{K+1} = \begin{cases} \max\{(1 - \zeta)\rho_K, \rho_{\min}\} & \text{if } \rho_K B_K^\eta \leq \max\{(1 - \zeta)\rho_K, \rho_{\min}\} , \\ \min\{(1 + \zeta)\rho_K, 1\} & \text{if } \min\{(1 + \zeta)\rho_K, 1\} \leq \rho_K B_K^\eta , \\ \rho_K B_K^\eta & \text{otherwise .} \end{cases} \quad (1.12)$$

Here ρ_K denotes the value of the density variable at iteration step K , and B_K is given by the expression

⁵ This term is in many circumstances called the *mutual energy density* – “mutual” since it involves the two fields u and \bar{u} . For compliance, $u = \bar{u}$, and the mutual energy density equals the strain energy density.

$$B_K = \Lambda_K^{-1} p \rho(x)^{p-1} E_{ijkl}^0 \varepsilon_{ij}(u_K) \varepsilon_{kl}(u_K), \quad (1.13)$$

where u_K is the displacement field at the iteration step K , determined from the equilibrium equation and dependent on ρ_K . Note that a (local) optimum is reached if $B_K = 1$ for densities ($\rho_{\min} < \rho < 1$). The update scheme (1.12) adds material to areas with a specific strain energy that is higher than Λ (that is, when $B_K > 1$) and removes it if the energy is below this value; this only takes place if the update does not violate the bounds on ρ . From integrating (1.11) one can see that Λ is proportional (by a factor p) to the average strain energy density of the part of the structure that is given by intermediate values of the density.

The variable η in (1.12) is a tuning parameter and ζ a move limit. Both η and ζ controls the changes that can happen at each iteration step and they can be made adjustable for efficiency of the method. Note that the update ρ_{K+1} depends on the present value of the Lagrange multiplier Λ , and thus Λ should be adjusted in an inner iteration loop in order to satisfy the active volume constraint. It is readily seen that the volume of the updated values of the densities is a continuous and decreasing function of the multiplier Λ . Moreover, the volume is strictly decreasing in the interesting intervals, where the bounds on the densities are not active in all points (elements of a FEM discretization). This means that we can uniquely determine the value of Λ , using a bisection method or a Newton method. The values of η and ζ are chosen by experiment, in order to obtain a suitable rapid and stable convergence of the iteration scheme. A typical useful value of η and ζ is 0.5 and 0.2, respectively.

It is noted above that the optimality criteria method is closely related to the concept of fully stressed design. However, it is important to note that the conditions (1.9, 1.10) only imply that the specific strain energy is constant in areas of *intermediate* density, while it is lower in regions with a density $\rho = \rho_{\min}$ and higher in regions with a density equal to 1.

The type of algorithm described above has been used to great effect in a large number of structural topology design studies and is well established as an effective (albeit heuristic) method for solving large scale problems [6], [7]. The effectiveness of the algorithm comes from the fact that each design variable is updated independently of the update of the other design variables, except for the rescaling that has to take place for satisfying the volume constraint. The algorithm can be generalized to quite a number of structural optimization settings (see for example Rozvany (1989), Rozvany (1992)), but it is not always straightforward. For cases where for example constraints of a non-structural nature should be considered (e.g., representing geometry considerations), when non-self-adjoint problems are considered or where physical intuition is limited, the use of a mathematical programming method can be a more direct way to obtain results. Typically, this will be computationally more costly, but a careful choice of algorithm can make this approach as efficient as the optimality criteria method (see section 1.2.3).

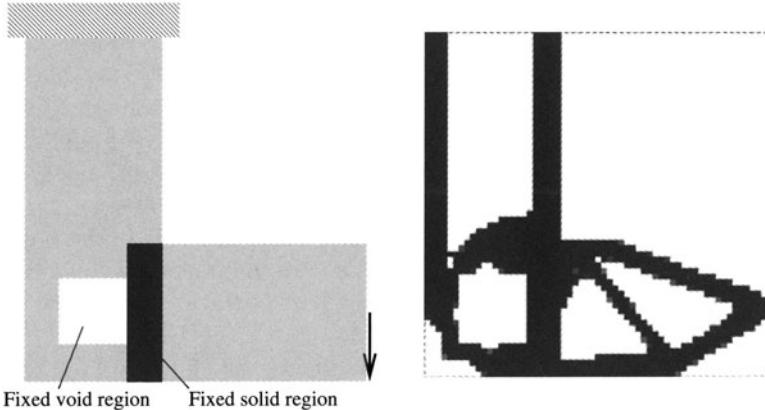


Fig. 1.4. The possibility of letting the design area be a sub-area of the reference domain. Same design domain as in Fig. 1.3 but with a square area fixed to be void and a rectangular area fixed to be solid. The compliance of the optimized topology is 25% lower than in Fig. 1.3 due to the restricted design domain.

1.2.2 Implementation of the optimality criteria method

In sections 1.1.2 and 1.2.1 we have outlined the basic ingredients of the optimality criteria method for implementing the material distribution procedure for topology design. These consist of the basic parametrization of design through the design-stiffness relation given by an appropriate interpolation scheme and the update scheme for the density based on the optimality conditions. Finally, this update scheme is based on the ability to solve the equilibrium equations, and here we presume this to be performed by the finite element method.

Computational procedure The direct method of topology design using the material distribution method is based on the numerical calculation of the globally optimal distribution of the density of material ρ which is the design variable. For an interpolation scheme that properly penalizes intermediate densities (cf., discussion in section 1.1.2) the resulting 0-1 (or black and white) design is actually the primary target of our scheme. The optimality criteria method for finding the optimal topology of a structure constructed from a single isotropic material then consists of the following steps:

Pre-processing of geometry and loading:

- Choose a suitable reference domain (the ground structure) that allows for the definition of surface tractions, fixed boundaries, etc. (see Fig. 1.4).
- Choose the parts of the reference domain that should be designed, and what parts of the ground structure that should be left as solid domains or voids (see Fig. 1.4).

- Construct a finite element mesh for the ground structure. This mesh should be fine enough in order to describe the structure in a reasonable resolution bit-map representation. Also, the mesh should make it possible to define the a priori given areas of the structure by assigning fixed design variables to such areas. The mesh is unchanged through-out the design process.
- Construct finite element spaces for the independent fields of displacements and the design variables.

Optimization:

Compute the optimal distribution over the reference domain of the design variable ρ . The optimization uses a displacement based finite element analysis and the optimality update criteria scheme for the density. The structure of the algorithm is:

- Make initial design, e.g., homogeneous distribution of material. The iterative part of the algorithm is then:
- For this distribution of density, compute by the finite element method the resulting displacements and strains.
- Compute the compliance of this design. If only marginal improvement (in compliance) over last design, stop the iterations. Else, continue. For detailed studies, stop when necessary conditions of optimality are satisfied.
- Compute the update of the density variable, based on the scheme shown in section 1.2.1. This step also consists of an inner iteration loop for finding the value of the Lagrange multiplier Λ for the volume constraint.
- Repeat the iteration loop.

For a case where there are parts of the structure which are fixed (as solid and/or void) the updating of the design variables should only be invoked for the areas of the ground structure which are being redesigned (reinforced).

Post-processing of results:

- Interpret the optimal distribution of material as defining a shape, for example in the sense of a CAD representation.

For the method above, one should at an initial stage decide on a choice of a basic interpolation scheme, for example SIMP. It is interesting to note that topology optimization using for example SIMP with a suitable high value of the power p gives rise to very well defined designs consisting almost entirely of areas with material or no material and very little area with intermediate density of material, i.e. very little grey.

It is important to underline that the algorithm just described can be implemented on any type of finite element mesh and any type of reference domain Ω (ground structure). This gives a significant flexibility to the method in terms of defining boundary conditions and non-design parts of the structure. Nevertheless, in many cases one works with rectangular (in 2-D) or

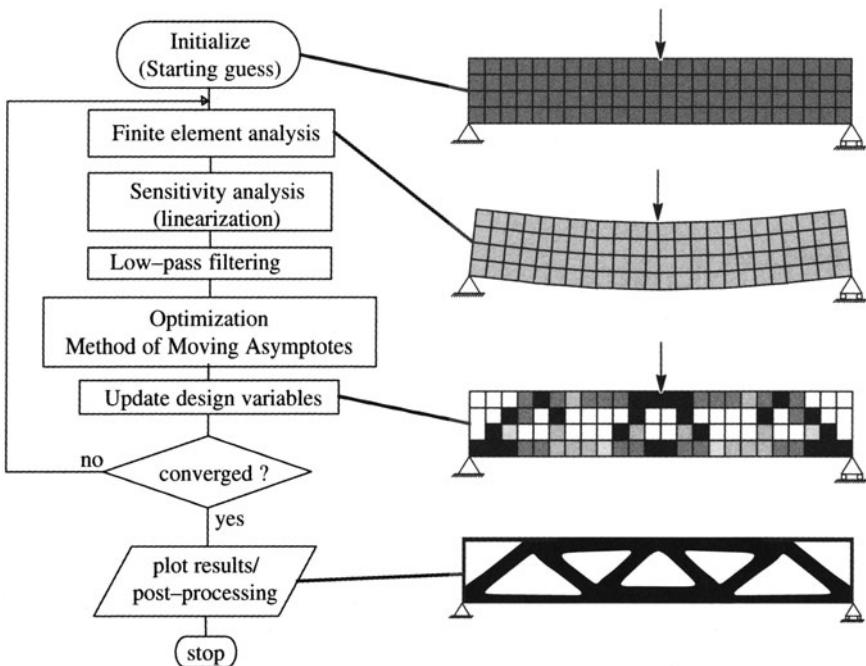


Fig. 1.5. The flow of computations for topology design using the material distribution method and the Method of Moving Asymptotes (MMA) for optimization. The low-pass filter step (filtering of sensitivities) is discussed in Sec. 1.3.1.

box-like (in 3-D) domains, and with a mesh consisting of squares or cubes. This simplifies implementation and can be employed to speed up the analysis part of the procedure, see section 1.2.4.

On programming complexity The procedure described above does not require any great programming efforts in order to solve the compliance topology design problem. When access to a FEM code is provided, only a few lines of extra code is required for the update scheme and for the computation of the energies involved. If for example a rectangular design domain is considered and one uses square elements and a Q4 interpolation of displacements and element wise constant densities, a complete program *including* FE analysis and plotting of the resulting designs can be written in 99 lines of Matlab code (see appendix 5.1.1). This actually also includes a filtering procedure that caters for the so-called checkerboard and mesh-dependency problems associated with our design problem (see section 1.3.2 for further details).

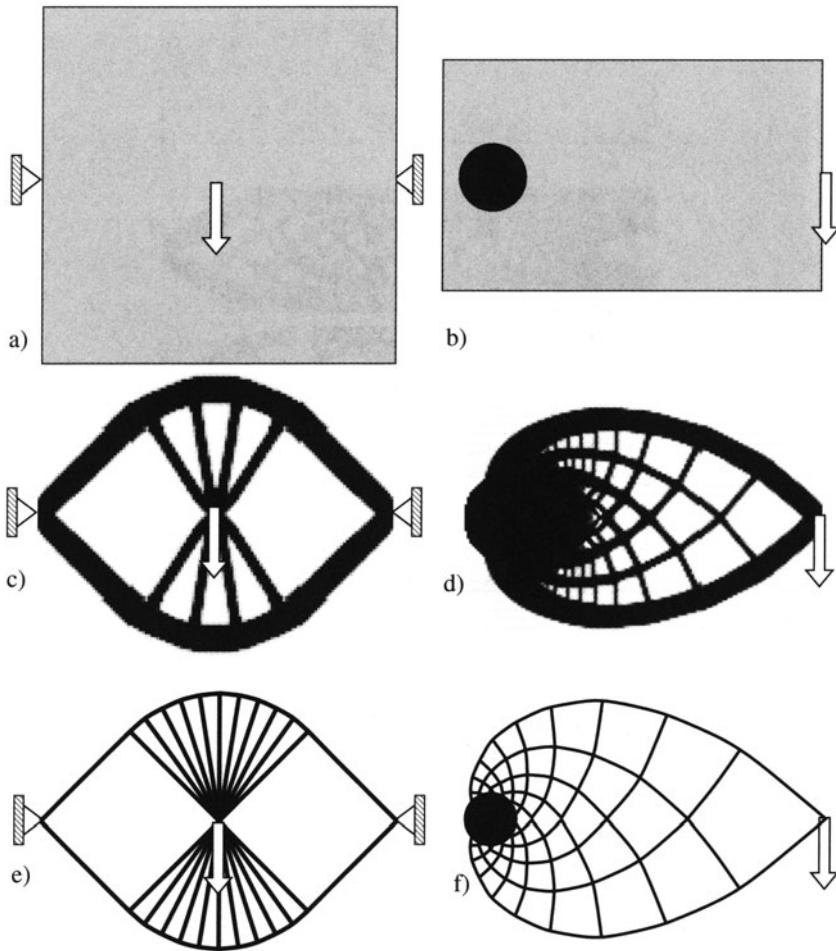


Fig. 1.6. Two examples of topology design for minimum compliance compared with optimal Michell type structures (Michell 1904). a) and b) design domains, c) and d) topology optimized solutions and e) and f) corresponding Michell type optimal solutions (from Sigmund 2000b).

1.2.3 Sensitivity analysis and mathematical programming methods

The use of mathematical programming algorithms for solving problems in structural optimization is well established and described in detail in the literature, for sizing as well as shape design problems [1]. The standard procedure is to consider the design problem as an optimization problem in the design variables *only*, and with the displacement field regarded as a function of these design variables. The displacement fields are given *implicitly* in terms of the

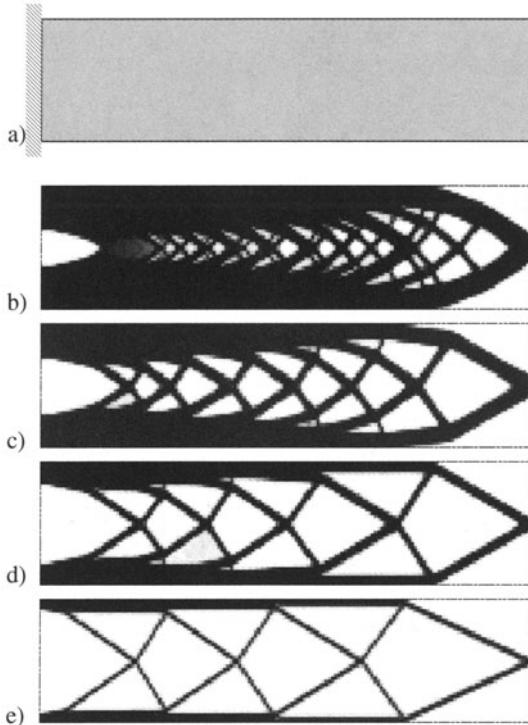


Fig. 1.7. The influence of volume fraction. A long cantilever beam discretized by 6400 square elements and optimized for volume fractions b) 80%, c) 60%, d) 40% and e) 20%. For low amounts of volume, truss-like structures are predicted.

design variables through the equilibrium equation and finding the derivatives of the displacements with respect to the design variables is termed *sensitivity analysis* [1].

The basic idea of the material distribution technique for topology design is to rephrase the problem as a sizing problem for the density ρ on a fixed domain. Thus the technique outlined above carries over to topology design as well. The major challenge, though, is to apply mathematical programming software that is well geared to cope with many variables and typically a moderate number of constraints. Here and in the following we will rely on the MMA algorithm, with “MMA” being the acronym for Method of Moving Asymptotes (Svanberg 1987, Svanberg 2002). This algorithm has proven itself to be versatile and well suited for large scale topology optimization problems.

Sensitivity analysis In order to complement the presentation of the optimality criteria method, we will here work with the FEM form of the minimum compliance problem, that is the problem (cf., (1.2)):

$$\begin{aligned}
& \min_{\mathbf{u}, \rho_e} \mathbf{f}^T \mathbf{u} \\
\text{s.t. : } & \left(\sum_{e=1}^N \rho_e^p \mathbf{K}_e \right) \mathbf{u} = \mathbf{f} , \\
& \sum_{e=1}^N v_e \rho_e \leq V, \quad 0 < \rho_{\min} \leq \rho_e \leq 1, \quad e = 1, \dots, N .
\end{aligned} \tag{1.14}$$

When solving this by a mathematical programming algorithm we first rewrite the problem as a problem in the design variables only:

$$\begin{aligned}
& \min_{\rho_e} c(\rho_e) \\
\text{s.t. : } & \sum_{e=1}^N v_e \rho_e \leq V, \quad 0 < \rho_{\min} \leq \rho_e \leq 1, \quad e = 1, \dots, N ,
\end{aligned} \tag{1.15}$$

where the equilibrium equation is considered as part of a function-call:

$$c(\rho_e) = \mathbf{f}^T \mathbf{u} , \quad \text{where } \mathbf{u} \text{ solves : } \sum_{e=1}^N \rho_e^p \mathbf{K}_e \mathbf{u} = \mathbf{f} . \tag{1.16}$$

When gradients are required by the optimization algorithm employed to solve (1.15), these are easily derived for the objectives and constraints involving only ρ . For functions that depend on the displacements also, derivatives can be obtained by the chain-rule. These expressions will then contain derivatives of the displacement, which in turn can be obtained by taking the derivative of the equilibrium equation $\mathbf{K}\mathbf{u} = \mathbf{f}$. In topology design we typically work with a moderate number of constraints, so the most effective method for calculating derivatives is to use the *adjoint* method, where the derivatives of the displacement are not calculated explicitly. For the minimum compliance problem (1.15) at hand we rewrite the function $c(\rho)$ by adding the zero function:

$$c(\rho) = \mathbf{f}^T \mathbf{u} - \tilde{\mathbf{u}}^T (\mathbf{K}\mathbf{u} - \mathbf{f}) ,$$

where $\tilde{\mathbf{u}}$ is any arbitrary, but *fixed* real vector. From this, after rearrangement of terms, we obtain that

$$\frac{\partial c}{\partial \rho_e} = (\mathbf{f}^T - \tilde{\mathbf{u}}^T \mathbf{K}) \frac{\partial \mathbf{u}}{\partial \rho_e} - \tilde{\mathbf{u}}^T \frac{\partial \mathbf{K}}{\partial \rho_e} \mathbf{u} .$$

This can in turn be written as

$$\frac{\partial c}{\partial \rho_e} = -\tilde{\mathbf{u}}^T \frac{\partial \mathbf{K}}{\partial \rho_e} \mathbf{u} ,$$

when $\tilde{\mathbf{u}}$ satisfies the adjoint equation:

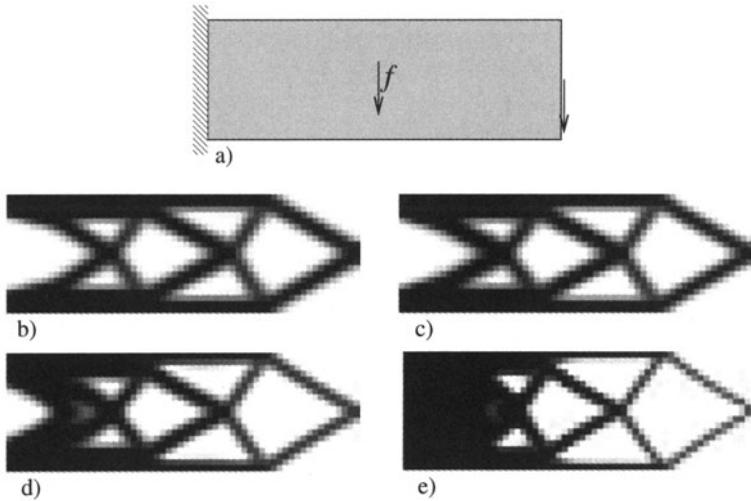


Fig. 1.8. Influence of self-weight on the topology optimized cantilever beam. Here, the load vector is design dependent (i.e. $\mathbf{f} = \mathbf{f}(\rho)$) and the sensitivity of compliance including self-weight can be found as $\frac{\partial c}{\partial \rho_e} = -\mathbf{u}_e^T \frac{\partial \mathbf{K}_e}{\partial \rho_e} \mathbf{u}_e + \frac{\partial \mathbf{f}^T}{\partial \rho_e} \mathbf{u}_e$. a) Design domain and loads, b) resulting topology for zero self weight, c) resulting topology for self weight equal to 1.2 times the non-structural load, d) resulting topology for self weight equal to 6 times the non-structural load and e) resulting topology for self weight equal to 24 times the non-structural load.

$$\mathbf{f}^T - \tilde{\mathbf{u}}^T \mathbf{K} = 0 .$$

This latter equation is in the form of an equilibrium equation and for compliance we see that we obtain directly that $\tilde{\mathbf{u}} = \mathbf{u}$ (normally the adjoint equation requires additional computations). Moreover, the form of the stiffness matrix means that the derivatives of the compliance $c(\rho)$ for problem (1.15) is of the particularly simple form:

$$\frac{\partial c}{\partial \rho_e} = -p \rho_e^{p-1} \mathbf{u}^T \mathbf{K}_e \mathbf{u} . \quad (1.17)$$

Thus derivatives for the minimum compliance problem are extremely easy to compute. Also, one notices that the derivative is “localized” in the sense that the derivative only involves information at the element level; however, there is an effect from other design variables hidden in the displacement \mathbf{u} . Finally, we see that the sensitivity is negative for all elements, so physical intuition is confirmed in that additional material in any element decreases compliance, that is, makes the structure stiffer.

The basics of MMA The Method of Moving Asymptotes (MMA) and its “mother” method CONLIN are mathematical programming algorithms

well suited for topology design⁶. They are in nature similar to methods like Sequential Linear Programming (SLP) and Sequential Quadratic Programming (SQP) for solving smooth, non-linear optimization problems, in the sense that they work with a sequence of simpler approximate subproblems of given type. For MMA and CONLIN these subproblems are separable and convex and are constructed based on sensitivity information at the current iteration point as well as some iteration history. At each iteration point this subproblem is solved by for example a dual method or by an interior point algorithm (primal-dual algorithm), and the solution to the subproblem is then used as the next design in the iterative procedure.

In MMA the approximation of a function F of n real variables $\mathbf{x} = (x_1, \dots, x_n)$ around a given iteration point \mathbf{x}^0 has the form

$$F(\mathbf{x}) \approx F(\mathbf{x}^0) + \sum_{i=1}^n \left(\frac{r_i}{U_i - x_i} + \frac{s_i}{x_i - L_i} \right),$$

where the numbers r_i, s_i are chosen as

$$\begin{aligned} \text{if } \frac{\partial F}{\partial x_i}(x^0) > 0 \text{ then } r_i &= (U_i - x_i^0)^2 \frac{\partial F}{\partial x_i}(x^0) \quad \text{and} \quad s_i = 0, \\ \text{if } \frac{\partial F}{\partial x_i}(x^0) < 0 \text{ then } r_i &= 0 \quad \text{and} \quad s_i = -(x_i^0 - L_i)^2 \frac{\partial F}{\partial x_i}(x^0), \end{aligned}$$

and where, loosely speaking, the positive numbers U_i, L_i control the range for which the approximation of F can generate reasonable answers for our optimization problem (the parameters U_i, L_i give vertical asymptotes for the approximations of F and is the source of the name of the algorithm). In the optimization algorithm, the values of U_i, L_i for each function of the problem are updated in each iteration, depending on the iteration history so far.

A central aspect of MMA and CONLIN is the use of such *separable* and convex approximations. The former property means that the necessary conditions of optimality of the subproblems do not couple the primary variables (the design variables) while the convexity means that dual methods or primal-dual methods can be used. Together this has an immense effect on reducing the computational effort needed to solve the subproblems, especially for problems with only a few constraints.

Experience over the last couple of years have shown that the convex approximation methods are very efficient for topology optimization. Typically in these problems one chooses to work with a large number of design variables (for the raster representation of the design, one operates with one or more density variables per element in the finite element mesh) and try to formulate the optimization problem with a fairly limited number of constraints. Compared to optimality criteria methods, the use of a mathematical programming tool does provide an added flexibility to topology design. One

⁶ The use of convex and separable approximations was first introduced with CONLIN (see, e.g. Fleury (1993) and references therein).

avoids the development and coding of new algorithms for each new problem that is to be solved, and it is also possible to handle geometry considerations and situations where physical intuition is limited.

We close this discussion by noting that for the minimum compliance problem the use of the optimality criteria method or the use of MMA in essence involves the same type of computations (see also Borrvall & Petersson (2001a)). We found in (1.17) that the sensitivity of compliance is negative for any element density ρ_e . Thus an MMA approximation of the compliance gives a subproblem after iteration step K in the form

$$\begin{aligned} \min_{\rho_e} & \left\{ c(\rho^K) - \sum_{e=1}^N \frac{(\rho_e^K - L_e)^2}{\rho_e - L_e} \frac{\partial c}{\partial \rho_e}(\rho^K) \right\} \\ \text{s.t. :} & \sum_{e=1}^N v_e \rho_e \leq V, \quad 0 < \rho_{\min} \leq \rho_e \leq 1, \quad e = 1, \dots, N. \end{aligned} \quad (1.18)$$

Solving this problem by a dual method now involves steps similar to the ones performed in section 1.2.1 for the optimality criteria method. First one minimizes the Lagrange functional

$$\mathcal{L} = c(\rho^K) - \sum_{e=1}^N \frac{(\rho_e^K - L_e)^2}{\rho_e - L_e} \frac{\partial c}{\partial \rho_e}(\rho^K) + \Lambda \left(\sum_{e=1}^N v_e \rho_e - V \right),$$

with respect to densities satisfying $\rho_{\min} \leq \rho_e \leq 1$, $e = 1, \dots, N$. Using convexity and that \mathcal{L} is separable, this optimization can easily be performed, element by element. For the case where $L_e = 0$ this results in exactly the optimality criteria update scheme given in (1.12), with move limit $\zeta = \infty$ and tuning parameter $\eta = 0.5$. The second step of the dual method is to maximize the resulting functional with respect to Λ , and as for the optimality criteria method this corresponds to adjusting the value of Λ so that the update scheme gives a density $\rho^{(K+1)}$ that satisfies the volume constraint. In the actual implementation of MMA, one chooses the asymptote parameters L_e more cleverly, improving speed of convergence.

In conclusion, we note that MMA is an invaluable tool for topology optimization problems. For simple compliance optimization problems it may be a bit slower than the OC method but for more complicated problems involving several constraints MMA stands for excellent convergence properties. Also, the advanced versions of the program caters for more complex problems formulations including min-max formulations etc. Contestants to MMA may be Sequential Linear Programming (SLP) method, CONLIN and other first order methods. Unless otherwise noted, we use either OC or MMA for all the applications shown in this book.

1.2.4 Implementation - the general concept

The use of mathematical programming techniques does not change the general flow of a topology design procedure. Thus, compared to the optimality criteria based method for topology design described in section 1.2.2, it only influences the optimization step of the scheme. This iterative loop becomes:

Optimization with, for example, MMA:

- Make initial design, e.g., homogeneous distribution of material. The iterative part of the algorithm is then:
- For this distribution of density, compute by the finite element method the resulting displacements.
- Compute the compliance of this design and the associated sensitivity with respect to design changes. If only marginal improvement (in compliance) over last design, stop the iterations. Else, continue.
- Compute the update of the density variable, based on the MMA approximate subproblem solved by a dual or a primal-dual method.
- Repeat the iteration loop.

The flow of computations sketched above shows the general concept, but programming details may be somewhat different, depending on how the specific mathematical programming software is structured. What is shown here is what is typically called an externally controlled optimization loop. In some cases an alternative structure is used, where the user provides subroutines that compute function values and sensitivity information, and the optimization software runs the loops “internally”.

For the compliance problem the use of for example MMA does not change significantly the programming effort required to implement the topology design procedure, and it can also be easily implemented in the 99 line Matlab code mentioned earlier (see appendix 5.1.1)⁷.

With this general outline of computations it is possible to generalize the concept to many other design settings of interest. This is the subject of Chap. 2.

A publicly available topology optimization software The topology optimization procedure as described in the preceding subsections has been implemented as an interactive design program made available to the public at the internet address <http://www.topopt.dtu.dk>. The program which is described in detail in Tcherniak & Sigmund (2001) solves a standard 3-load case compliance optimization problem. The topology optimization code is written in FORTRAN77 and the Graphical User Interface (GUI) and the

⁷ The MMA-code is available both in FORTRAN and MATLAB versions from K. Svanberg. The codes are free for academic use.

file-transfer system is written in JAVA. The structures are independently of aspect ratio always discretized by approximately 1000 elements⁸.

Screen-dumps of the Graphical User Interface of the TOPOPT program is shown in Fig. 1.9 together with a plot of the output. By pressing the left panel buttons and using the mouse and the delete key, the user can define the design domain, passive and active areas, load-cases and supports. After choosing the volume fraction and hitting the submit button, a sequence of GIF-files illustrating the iteration history appears on the screen. Depending on network load and connection, the solution of the problem may take from 10 seconds and up to a minute. As an extra feature, the animation sequences may be downloaded after running the program.

Implementation details The apparent simplicity of computing minimum compliance optimal topologies is somewhat betrayed by certain details that one has to cater for if a generally useful and applicable tool is to be constructed. For example, as already mentioned, the scheme described till now needs to be supplemented with some computational device that controls the range of allowable density distributions, especially the so-called checkerboard patterns that are numerical artifacts related to the discretization of displacements and densities. Also, the problem is inherently mesh-dependent as formulated and will generally not converge with mesh-refinement. Rather, finer and finer structure of the designs will arise. These aspects are the subject of the following section 1.3.

Another topic which is relevant to discuss here is the effect of the power p of the SIMP interpolation. As mentioned, to obtain black and white (0-1) topologies one has to work with a fairly large value of p , say, above 3. Moreover, to eventually be able to interpret appearance of grey in the final design as composite material, one also requires p be at least 3 (cf., (1.5), and the discussion in 1.5.4). However, in implementations it is often seen that a too severe penalization of the intermediate density can lead to designs which are local minima and which are very sensitive to choice of the initial design for the iterative optimization procedure (one “jumps” too fast to a 0-1 design). Thus a continuation method is often advisable, which means that the power p is slowly raised through the computations, until the final design is arrived at (for a power satisfying our requirements). The scheme is not guaranteed to give a fully 0-1 design (see Stolpe & Svanberg (2001b)). Nonetheless, it works well in most cases, especially when combined with a filtering of sensitivities as described in section 1.3.1. It is advisable to use such a continuation scheme as the standard procedure.

It is here important to underline that for the minimum compliance problem the by far most time-consuming part of the computations is spent on solving the equilibrium equations (Borrvall & Petersson (2001a) report this

⁸ The current version of the web-program has been updated and additional features added by Lars Voxen Hansen and Thomas Buhl (DTU) in the spring of 2002.

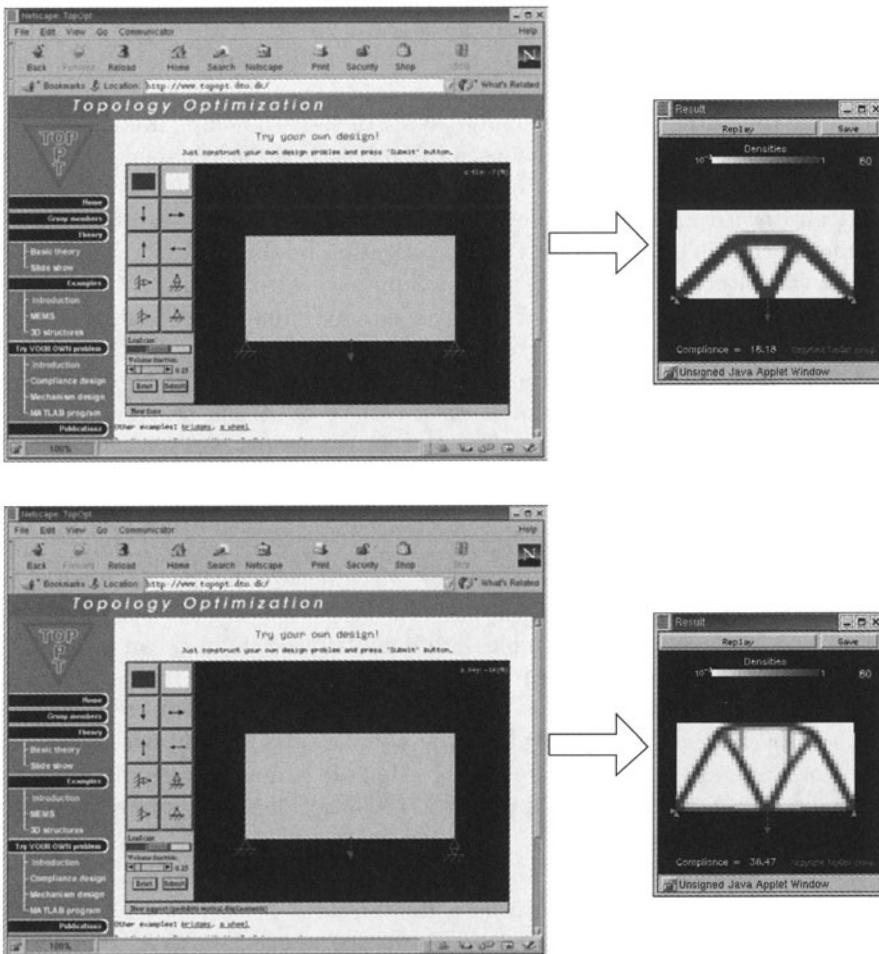


Fig. 1.9. A web-based topology optimization program (<http://www.topopt.dtu.dk>) for stiffness design of arbitrary two dimensional structures. The pictures demonstrate an investigation of the influence of boundary conditions on the optimal topology (from Tcherniak & Sigmund 2001).

share as up to 97%, in a parallel implementation). Thus it is critical for large problems, especially in 3-D, to improve on the efficiency of the analysis capability. Here the utilization of homogeneous meshes on rectangular or box-like domains is useful, as it removes the necessity for the repeated computation of local stiffness matrices. Also, the use of iterative solvers is useful in large scale problems, and may be required for storage reasons. The ultimate tool is to use vector computations and parallel computing, as for example reported in Borrval & Petersson (2001a), where also the MMA-based optimization is

parallelized - solutions to 3-D problems with up to 220.000 brick elements have been obtained (c.f., Figs. 1.11 and 1.12). Solutions to similar-sized problems have also been reported in DeRose Jr. & Díaz (2000) (c.f. Fig. 1.10), but here the paradigm for the analysis part of the problem is fundamentally changed. Instead of finite elements, a mesh-less, fictitious domain method is used, based on using wavelets in a Galerkin scheme. The advantage here is that an iterative, preconditioned conjugate gradient scheme exhibits a performance that is insensitive to the discretization level.

In the future we may also see new approaches to solving the topology optimization problem, based for example on developments in computational mechanics. One possibility is to solve the original combined optimization problem in one optimization routine and break with the tradition of viewing the equilibrium statement as a function call. This means that analysis becomes part of the iterative procedure and for example a Newton scheme will solve simultaneously the necessary conditions of optimality for the density as well as for the displacements (the latter being the equilibrium equations). This is what has been named a SAND approach (Simultaneous ANalysis and Design). It has for example been tested for topology design in a multigrid framework in Maar & Schulz (2000) and in the setting of free material design and truss design where for example semi-definite programming can be applied due to special problem structure (see Sects. 4.3 and 5.5). From a practical point of view this approach has the weakness that a premature termination of the procedure does not provide any analysis results for the current design, and the method also requires full integration of optimal design and analysis in one software system, i.e., optimization will no longer be an “add-on” to existing FE software.

1.2.5 Topology optimization as a design tool

In the following we will try to illustrate some basic features of the material distribution method when used for design, dealing for the moment only with compliance design. In a later section 1.4 of this chapter we shall describe the possible use of the topology design method as a pre-processor in an integrated design process where boundary variation techniques are employed for refining a design created by the topology design method.

In Chap. 2 we shall see that the topology design methodology over the last decade has matured immensely and that one can today cater for a broad range of structural objectives and constraints. Also, aspects of controlling geometric features can be handled, as illustrated in section 1.3.1. Moreover, a broad range of physics can be included in the modelling. This combined with progress in algorithms and computational power means that the topology design methodology is today much closer to being able to provide the user with a final design than just a few years ago, and for example in the design and manufacturing of MEMS (MicroElectroMechanical Systems, see section 2.6) case studies have shown that one can go directly from the topology

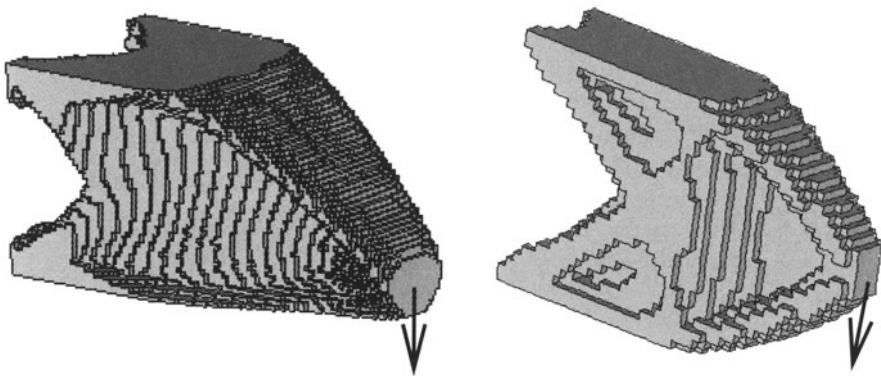


Fig. 1.10. Optimized 3D cantilever beams. Left: discretization by 221.184 design elements in a fictitious domain consisting of $128^3 = 1.097.152$ voxels. By courtesy of De Rose and A. Díaz (from DeRose Jr. & Díaz 2000).

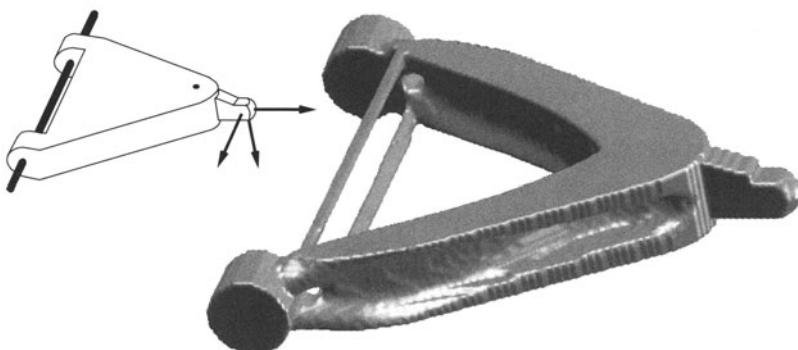


Fig. 1.11. Optimal design of a wishbone. Multiple load design using a very fine FE model (260.000 elements). Note that post-processing is basically not needed. By courtesy of T. Borrval.

design output to manufacturing. In many other circumstances the practical use of topology design is often at the level of a creative sparring partner in the *initial* phase of a design process. Thus the output of the topology design method is used to identify potentially good designs, the completion of the design being based on traditional skills of the design office. One effect of the topology method that cannot be underestimated is the efficient testing of the appropriateness of the model of loads and supports. As the *topology* is very sensitive to a proper modelling of the load environment, one can immediately discover discrepancies or inaccuracies in this modelling.

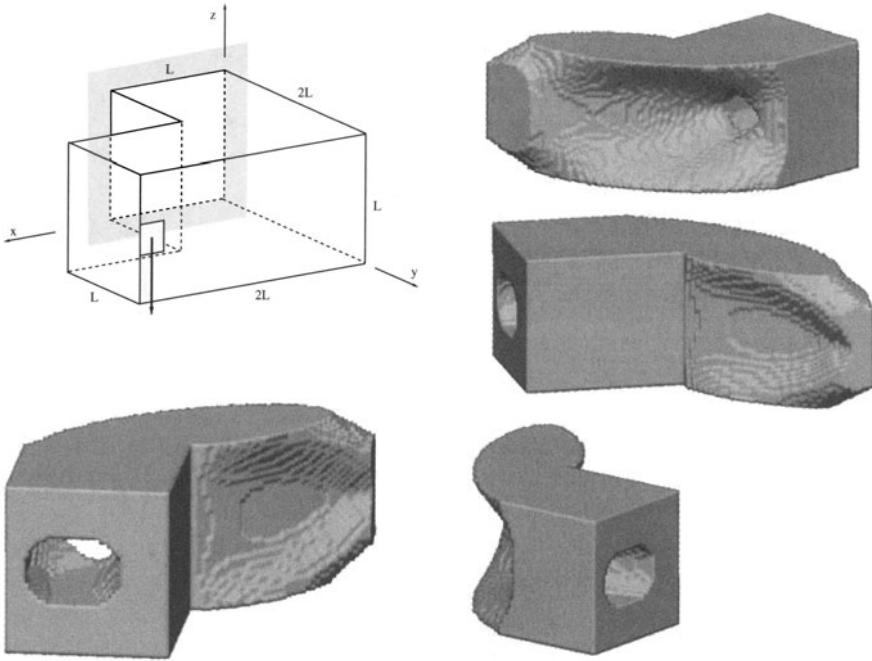


Fig. 1.12. Optimal design of cantilevered torsional beam in 3D. Design domain and different views of the optimized topology. Discretized using 128000 elements for one half of the problem and solved using parallel processing. By courtesy of T. Borrvall and J. Petersson (from Borrvall & Petersson 2001a).

Examples of topology design The material distribution method for topology design has been tested on a large number of examples, a few being illustrated in this and the preceding sections. The method allows for an efficient prediction of the optimal topology, the optimal shape and the optimal use of the prescribed possible support conditions. Also, it has proven to be a flexible and reliable design tool. The methodology has over the last decade become a fairly widespread tool in industrial applications, especially among some major car manufacturers (see section 2.12.4), and the appearance of commercial software has had an immense impact on the utilization of topology design methods in practise [23].

For an efficient use of topology design, the problem should be formulated on a ground structure (a reference domain) that is chosen as simple as possible to reduce the size of the analysis problem. The domain should, as described in section 1.2.2, allow for definition of loads and tractions and of boundary condition. The use of an automatic mesh generator will, of course, simplify the treatment of problems with complicated geometry such as non-simply connected reference domains. Complicated reference domains are needed for

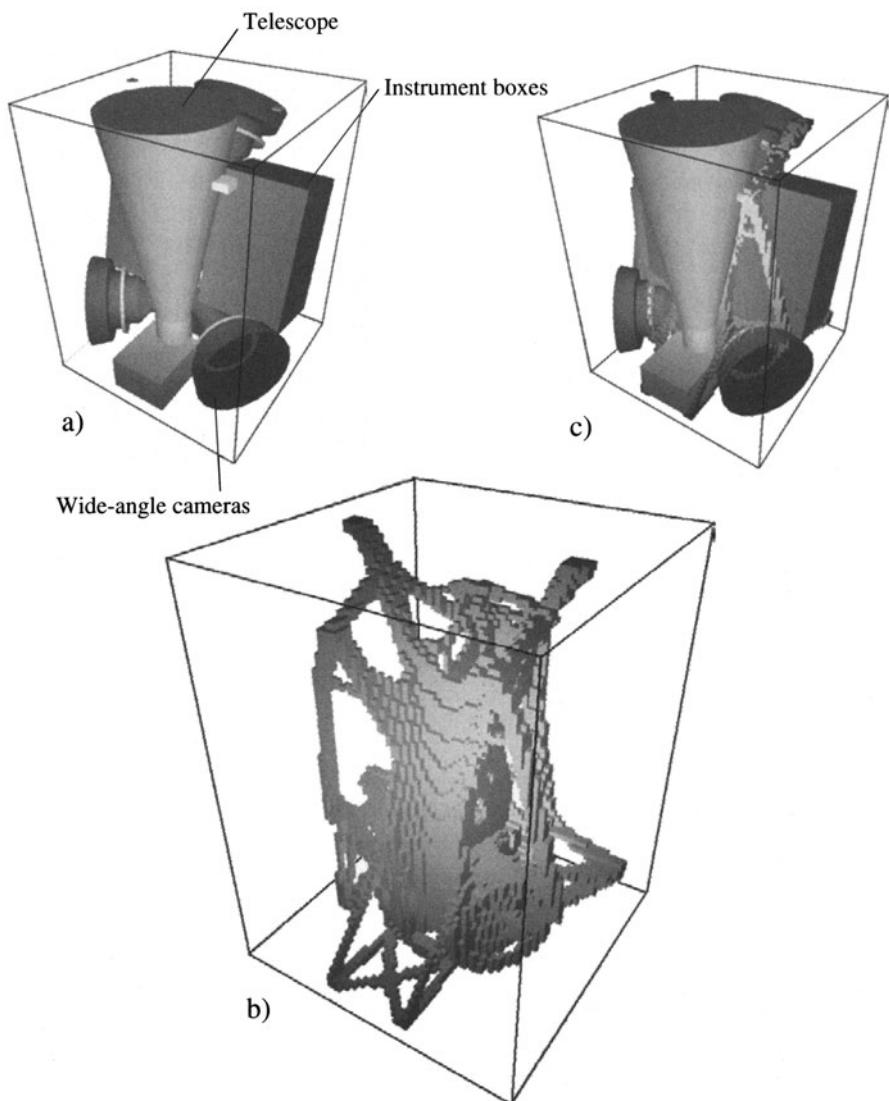


Fig. 1.13. Optimal design of frame for a satellite. Multiple load design using a 3-D model. The structure is discretized by 288.000 cubic finite elements. a) Design domain showing instruments around which the frame is to be designed. b) Optimized topology and c) optimized topology with instruments (from Sigmund 2000b).

cases where design requirements imply the exclusion of certain parts of space as parts of the structure. If the precise shapes of inner holes in a non-simply connected reference domain are unimportant, it is advisable to cater for such holes by fixing the density of material to be zero for the elements defining the

hole (or parts of it). These considerations have led to most examples being treated in rectangular reference domains, but the use of the method is of course not restricted to such domains.

For very low volume fractions, very fine discretization meshes are required when dealing with 0-1 designs, as the structures break up if coarse meshes are used. However, for high volume fractions, even coarse meshes give a very good indication of shape and topology and a good estimate of the optimal compliance. Note that for comparatively small volume fractions, the method predicts the lay-out of truss like structures and Michell frames (Fig. 1.2.2), thus supplementing lay-out theory and truss topology methods (see Chap. 4) for cases with a large, dense set of nodes; the material distribution method not only predicts the optimal connectivities, but also the optimal location of nodal points.

The application of interpolation schemes like SIMP that penalizes the designs to become of a 0-1 nature results in what one could classify as classically useful structures. One may argue that with present day technology for producing advanced composite materials one should certainly *not* remain limited by a wish to predict black and white designs only, but composites should also be part of the “structural universe”. This is the theme of Chap. 3.

Here and in the remainder of the monograph we show only a restricted number of examples of optimal topologies. Many more examples of topology designs from academia and industry can be found in the literature and on the web.

1.3 Complications

In the following we will discuss two important issues that significantly influences the computational results that can be obtained with the material distribution based topology design procedure. These are the appearance of checkerboards and the mesh-dependency of results. The former refers to the formation of regions of alternating solid and void elements ordered in a checkerboard like fashion and is related to the discretization of the original continuous problem. Mesh-dependence concerns the effect that qualitatively different optimal solutions are reached for different mesh-sizes or discretizations and this problem is rooted in the issue of existence of solutions to the continuous problem.

1.3.1 Mesh-refinement and existence of solutions

It is well-established that the 0-1 and the SIMP topology optimization problem formulated in section 1.1 lacks existence of solutions in its general con-

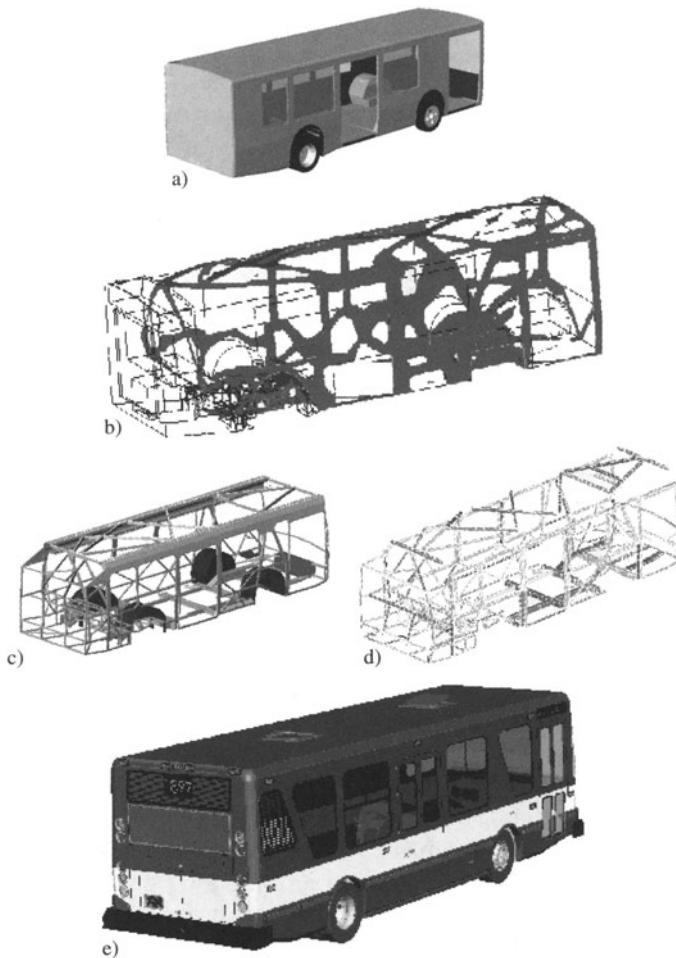


Fig. 1.14. Design of a lightweight city bus from preliminary design to final structural design. This example is courtesy of Altair Engineering, Inc., and has appeared in the Altair OptiStruct users manual and in (Thomas et al. 2002). At first, topology optimization is used to generate the optimum structural lay-out concept. The design space is shown in a) as the gray colored panels. b) The results of the topology optimization. c) CAD representation of the interpretation of the results of the topology optimization. d) Sizing optimization is performed on the hollow rectangular bar members of the bus structure. e) Final bus design. The shape of the windows was decided by the results of the structural needs identified by the topology optimization.

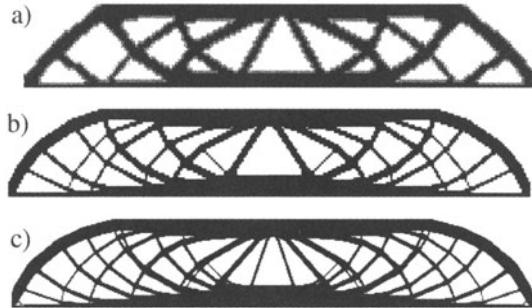


Fig. 1.15. Dependence of the optimal topology on mesh refinement for the MBB-beam example. Solution for a discretization with a) 2700, c) 4800 and d) 17200 elements.

tinuum setting⁹. The reason is that the introduction of more holes, without changing the structural volume, will generally increase the efficiency of a given structure. In the limit of this process one obtains structural variations in the form of microstructures that have an improved use of the material. Such microstructures are typically not isotropic and cannot be represented within the original design description of only isotropic material; one says that there is a lack of closedness of the admissible set of designs. In computational implementations this effect is seen as a numerical instability where a larger number of holes appear when a finer finite element mesh is employed. That is, refining the finite element mesh for the reference domain ultimately leads to a generation of a fine-scale internal structural lay-out similar in nature to the microstructures that theory predicts. Thus the *non-existence* of solutions is indeed a problem for the numerical solutions of the topology optimization problem. This dependence of the solutions on mesh-refinement is illustrated in figure 1.15, where an improved finite element discretization results in a much more detailed structure. Ideally, mesh-refinement should result in a better finite element modelling of the same optimal structure and a better description of boundaries – not in a more detailed and qualitatively different structure. As we shall show, there are actually efficient and uncomplicated ways to achieve mesh-independent procedures for obtaining 0-1 designs, so there is no reason to accept results that are inherently mesh-dependent.

The approach to generate macroscopic and mesh-independent 0-1 solutions is to reduce the space of admissible designs by some sort of global or local restriction on the variation of density, thus effectively ruling out the possibility for fine scale structures to formate. The techniques that have been suggested for enforcing such a restriction fall into three generic classes of methods. These consists of either adding constraints to the optimization problem, reducing directly the parameter space for the designs, or applying filters in the optimization implementation. For most of these methods, exis-

⁹ In any discretized version of the 0-1 problem, existence is trivial, as one has a design space with finitely many different design options.

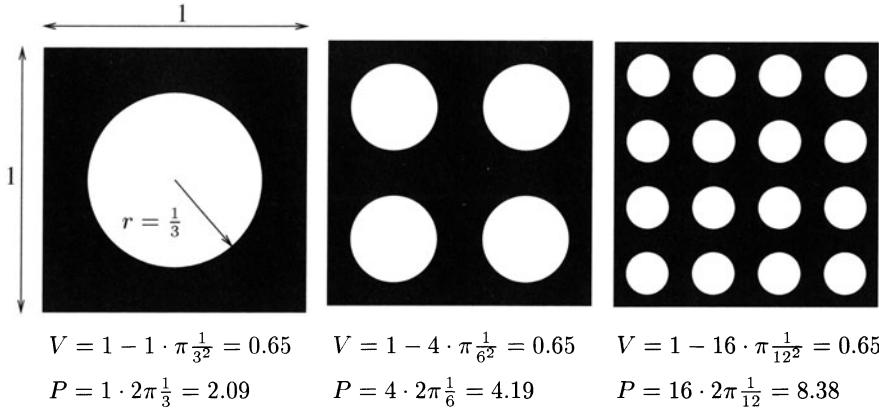


Fig. 1.16. An example of how smaller holes increase the perimeter, for a fixed volume. V is the volume and P is the perimeter of internal holes.

tence of solutions and also convergence of the FE approximations have been proved, providing a solid foundation [8].

We close this brief discussion by noting that the alternative to a restriction of the design space is to extend the space by allowing composites as admissible designs (see Chap. 3). For minimum compliance this lives up to our requirement of independence of mesh refinement, but also gives designs with large areas of “grey”. This is thus not an option if 0-1 designs are the goal¹⁰.

Perimeter control The *perimeter* of a mechanical element Ω^{mat} is, vaguely speaking, the sum of the lengths/areas of all inner and outer boundaries. Constraining the perimeter clearly limits the number of holes that can appear in the domain, (cf. figure 1.16) and existence of solutions to the perimeter controlled topology optimization is actually assured for both the discrete 0-1 setting and the interpolated version using SIMP. Also, it has been implemented for both situations and for 2-D as well as 3-D problems [8]. For the SIMP method, one can impose a constraint that mimics such a perimeter bound in the form of an upper bound on the *total variation*, $TV(\rho)$, of the density ρ . In case the function ρ is smooth, the total variation constraint is a L^1 bound on its gradient:

$$TV(\rho) = \int_{\mathbf{R}^n} \|\nabla \rho\| dx \leq P^*. \quad (1.19)$$

¹⁰In order to force the composite solutions to a 0-1 design, an explicit penalty on intermediate densities can be added - but this destroys the existence as one reverts to the ill-posed 0-1 problem.

For a 0-1 design, the total variation of ρ coincides with the perimeter of Ω^{mat} when ρ is 1 in Ω^{mat} and 0 elsewhere (in $\mathbf{R}^n, n = 2(3)$). In this case the constraint is expressed as

$$TV(\rho) = \sup \left\{ \int_{\mathbf{R}^n} \rho \operatorname{div} \varphi dx \mid \varphi \in C_c^1(\mathbf{R}^n, \mathbf{R}^n), \|\varphi\| \leq 1 \right\} \leq P^*, \quad (1.20)$$

where $C_c^1(\mathbf{R}^n, \mathbf{R}^n)$ denotes compactly supported vector valued C^1 functions.

For an element wise constant finite element discretization of the density the total variation can in 2-D be calculated as

$$P = \sum_{k=1}^K l_k \left(\sqrt{\langle \rho \rangle_k^2 + \epsilon^2} - \epsilon \right), \quad (1.21)$$

where $\langle \rho \rangle_k$ is the jump of material density through element interface k of length l_k and K is the number of element interfaces (here one should also count interfaces at the boundary of the domain Ω – else there will be bias towards having material at the borders of Ω). The parameter ϵ is a small positive number which is used to convert the non-differentiable absolute value into a differentiable term. This expression is exactly the total variation of the element-wise constant density when $\epsilon = 0$.

It should be mentioned that there is an inherent problem of assuring isotropy in an implementation of a discretized perimeter measure (this effect is also known from work in image segmentation, cf., Chambolle (1995)). Thus, in a regular 2-D mesh of squares, a bound on the discretized expression (1.21) will tend to favour structural edges parallel to those of the finite element mesh. This is caused by the effect that a straight edge of length 1 that is angled 45 degrees to the directions of the finite element mesh will be approximated by a jagged edge that has the perimeter $\sqrt{2}$. In contrast, the same edge has perimeter 1 when it is parallel to the mesh directions. In the limit of mesh refinement for a FE-mesh directed along the x_i -axes, the discretized perimeter measure (1.21) is thus rather the proper discretization of what is referred to a “taxi-cab” perimeter measure(cf., Petersson (1999b)):

$$TV_{\text{taxicab}}(\rho) = \int_{\mathbf{R}^2} \left(\left| \frac{\partial \rho}{\partial x_1} \right| + \left| \frac{\partial \rho}{\partial x_2} \right| \right) dx. \quad (1.22)$$

This means that the numerical results will approach solutions of a continuum topology optimization problem statement that includes a “perimeter bound” that actually measures the “length” of the boundary of the structure by projecting this onto the coordinate axes. This in turn implies that even though the perimeter constraint (1.21) assures convergence with respect to mesh refinement, a dependence on the choice of mesh will nonetheless be seen. This effect, however, does not change with mesh refinement. This directional bias of the results can be reduced considerably by considering more involved discrete versions of the perimeter measure, see Petersson, Beckers & Duysinx

(2000) and Borrvall (2001); in the latter reference more refined discretization schemes are also discussed.

The perimeter bound adds one extra constraint to the topology optimization problem and thus does not create any substantial problems for the use of an algorithm like MMA. However, it has been reported that the perimeter constraint can be quite difficult to approximate resulting in fluctuations in the design variables (this relates to the choice of the asymptotes of MMA). However, this can be solved by an internal loop procedure for the perimeter approximation which is computationally inexpensive compared to the equilibrium analysis (see Duysinx (1997)). Finally, one should note that choosing the bounding value of the perimeter constraint can be rather tricky, see below.

Other methods of restricting gradients One can also consider other types of gradient constraints for the SIMP method that ensure existence of solutions and convergence with mesh-refinement. These presuppose that ρ is sufficiently smooth for the bound to make sense and do not seem to have any equivalent for the discrete-valued 0-1 setting, in contrast to the perimeter measure discussed above.

One possibility is to constrain the local density variation by imposing pointwise bounds on the derivatives of the function ρ :

$$\left| \frac{\partial \rho}{\partial x_i} \right| \leq G, \quad (i = 1, 2, (3)). \quad (1.23)$$

This scheme, which in essence constrains the L^∞ norm of the gradient of ρ , does assure existence of solutions and convergence of the finite element scheme (Petersson & Sigmund 1998). The advantage of this gradient constraint is that it gives a well-defined local length scale. The constraint in (1.23) implies that a transition from void to void through full material has to take place over a distance that is longer than $2/G$, which is thus the width of the thinnest features of a feasible design. Unfortunately, an implementation results in a huge number of extra constraints in the optimization problem and the method must therefore be considered to be too slow for practical design problems, if implemented directly as constraints. However, if one approximates the L^∞ constraint (1.23) by a L^q constraint for suitable large q one can alternatively operate with just one global constraint (but choosing the constraint value can be tricky and requires experimentation for each design case) (see Borrvall (2001)).

The basic concept of a slope constraint can also be enforced by an adaptive constraint strategy in the optimization algorithm that is similar to adding move limits (Zhou, Shyy & Thomas 2001). This means that one only works with the values of the box-constraints on the density ρ , which at the $(K+1)$ -th iteration step are modified to restrict the variations in the design

$$\rho_i^{K+1} \geq \max\{\rho_{\min}, \rho_{j(i)}^K - D_{i, j(i)}G\}. \quad (1.24)$$

Here $j(i)$ is the element number of the element with the highest density value among all elements adjacent to the element i at the prior iteration step, and $D_{i, j(i)}$ is the distance between the centers of the elements i and $j(i)$. This strategy does not add to the computational complexity of the optimization procedure. However, it does require that the applied optimization algorithm can handle (temporary) violations of the box constraints. Furthermore, it is unclear whether “playing” with the move-limits will jeopardize convergence of the algorithm.

Another option is to implement a “global gradient constraint” by which we mean the norm of the function ρ in the Sobolov space $H^1(\Omega)$:

$$\|\rho\|_{H^1} = \left(\int_{\Omega} (\rho^2 + \|\nabla \rho\|^2) \, d\Omega \right)^{\frac{1}{2}} \leq M. \quad (1.25)$$

Proof of existence when including this bound in the minimum compliance problem can be found in appendix 5.2.2 (for three dimensional problems the proof requires that the power in SIMP satisfies $p < 3$). Note that we for any finite element discretization of the ground structure Ω can choose a large enough bound M on the norm of ρ so that the norm constraint remains inactive, thus seemingly returning to the original formulation for this discretization. Thus implementation of (1.25) also requires utmost care and should involve experimenting with a range of values of the bound M . A global gradient constraint can also be formulated with the term ρ^2 removed from (1.25), so that the constraint becomes a L^2 constraint on the gradient of ρ . Numerical experiments with global gradients in the setting of topology optimization can be found in Borrvall (2001), where also L^q constraints in general are considered.

Filtering the density The techniques above impose explicit limitations on the allowable density distributions that can appear in the optimal design, and as such have to be catered for as constraints in the optimization formulation. An alternative to this is to directly limit the variations of the densities that appear in the set of admissible stiffness tensor E_{ad} by only admitting *filtered* densities in the stiffness [8]. Thus the SIMP method is modified to the following reduced design space:

$$\begin{aligned} E_{ijkl}(x) &= ((\rho * K)(x))^p E_{ijkl}^0, \quad \rho \in L^\infty(\Omega), \\ (\rho * K)(x) &= \frac{1}{\langle K \rangle} \int_{\Omega} \rho(y) K(x - y) \, dy, \quad \langle K \rangle = \int_{R^n} K(y) \, dy, \\ \int_{\Omega} \rho(x) \, d\Omega &\leq V; \quad 0 \leq \rho(x) \leq 1, \quad x \in \Omega, \end{aligned} \quad (1.26)$$

where K is a convolution kernel, for example

$$K_r(x) = \begin{cases} 1 - \frac{\|x\|}{r} & \text{if } \|x\| \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

The filter radius r is fixed in the formulation and implies the enforcement of a fixed length-scale in the stiffness distribution. The filtering means that the stiffness in a point x depends on the density $\rho(x)$ in all points of a neighborhood of x . This implies a smoothing of the stiffness fields in a fashion similar to a filtering of an image. The smoothing and the fixed scale means that this method gives existence of solutions and convergence with refinement of the FE mesh. Loosely speaking, the reason that the filter removes any fine scale behaviour of the density ρ is that such variations in the mechanical analysis (via the filtering $(\rho * K)$) appears like a grey which is penalized by SIMP. Generally this method results in density fields ρ that are bi-valued, but the stiffness distribution $(\rho * K)^p$ is more “blurred” with grey boundaries. In a sense this is an ambiguity, as the mechanical analysis is done on the filtered density¹¹. For implementation, the differences compared to the standard procedure described in section 1.2 are that the element stiffness matrices in the finite element analysis are defined by weighted averages of the stiffnesses of neighbouring elements, and the sensitivity information should be modified to cater for the redefined stiffness tensor (this means that the sensitivity of the compliance with respect to $\rho(x)$ will involve the mutual energy of a neighborhood of x).

Filtering the sensitivities Computational experience has shown that filtering of the sensitivity information of the optimization problem is a highly efficient way to ensure mesh-independency [8]. This means modifying the design sensitivity of a specific element, based on a weighted average of the element sensitivities in a *fixed* neighborhood. Such a filter is purely heuristic but it produces results very similar to for example those obtained by a local gradient constraint, it requires little extra CPU-time and it is very simple to implement as it does not add to the complexity of the optimization problem (no extra constraints need to be considered). Similar ideas of weighted averages have been used to ensure mesh-independence for simulations of bone-remodelling and for analysis with plastic-softening materials (Mullender, Huiskes & Weinans. 1994, Leblond, Perrin & Deveaux 1994).

The scheme works by modifying the element sensitivities of the compliance as follows:

$$\widehat{\frac{\partial f}{\partial \rho_k}} = \frac{1}{\rho_k \sum_{i=1}^N \hat{H}_i} \sum_{i=1}^N \hat{H}_i \rho_i \frac{\partial f}{\partial \rho_i}, \quad (1.27)$$

¹¹ In contrast, a non-filtered version of SIMP evaluates a 0-1 design with the same distribution of stiffness as represented by the density.

where N is the total number of elements in the mesh and where the *mesh-independent* convolution operator (weight factor) \hat{H}_i is written as

$$\hat{H}_i = r_{\min} - \text{dist}(k, i), \quad \{i \in N \mid \text{dist}(k, i) \leq r_{\min}\}, \quad k = 1, \dots, N. \quad (1.28)$$

In this expression, the operator $\text{dist}(k, i)$ is defined as the distance between the center of the element k and the center of an element i . The convolution operator \hat{H}_i is zero outside the filter area. The convolution operator for element i is seen to decay linearly with the distance from element k . It is worthwhile noting that the sensitivity (1.27) converges to the original sensitivity when the filter radius r_{\min} approaches zero and that all sensitivities will be equal (resulting in an even distribution of material) when r_{\min} approaches infinity. This filter is implemented in the Matlab code of appendix 5.1.1.

Unfortunately, the theoretical basis for the method is not yet understood. Also, it is unclear exactly what problem we are solving. However, numerous applications, many of which are shown in this monograph are based on this filtering method. It has been applied to 2 and 3 dimensional problems, to problems with up to 20 structural or other constraints, to problems involving multiple areas of physics and it has been an invaluable tool in designing extremal material structures (c.f., Sect. 2.10). Furthermore, it gives results that are stable under mesh-refinement and maintain a minimum length-scale that is controlled by the filter radius r_{\min} . Also, experience shows that the filter somehow improves the computational behaviour of the topology design procedures as it delays the tendency of the SIMP scheme to get “stuck” in 0-1 designs (this is discussed in more detail in Chap. 2). Fig. 1.17 shows an example of mesh-independent designs obtain by the filter. In the following we will refer to this filtering as the *mesh-independence filter*.

Monotonicity based length scale control In some recent work a scheme called the MOLE method (*MONotonicity based minimum LEnghth scale*) has been proposed for the control of length-scale (Poulsen 2003). As for the perimeter and global gradient control it adds one extra constraint to the optimization problem, but in this case the non-negative constraint function should have value *zero* for the minimum length scale restriction to be satisfied. Moreover, one can explicitly specify the desired minimum width d of material parts and void inclusions. Thus it provides similar exact control as when using local gradients, but within just one constraint.

The idea is to pass a circular “filter” window over the design and measure if the density ρ along four equally spaced diagonals¹² is monotonic or not. The reason for testing over four diagonals is that a test only along the horizontal and vertical directions would not be able to detect the fine-scale variation of a corner to corner “hinge” in a regular mesh, while testing along the diagonals only would prevent checkerboard detection. On the other hand testing along more directions would make almost any design infeasible (see below).

¹² Horizontal, vertical and at $\pm\pi/4$ from the horizontal.

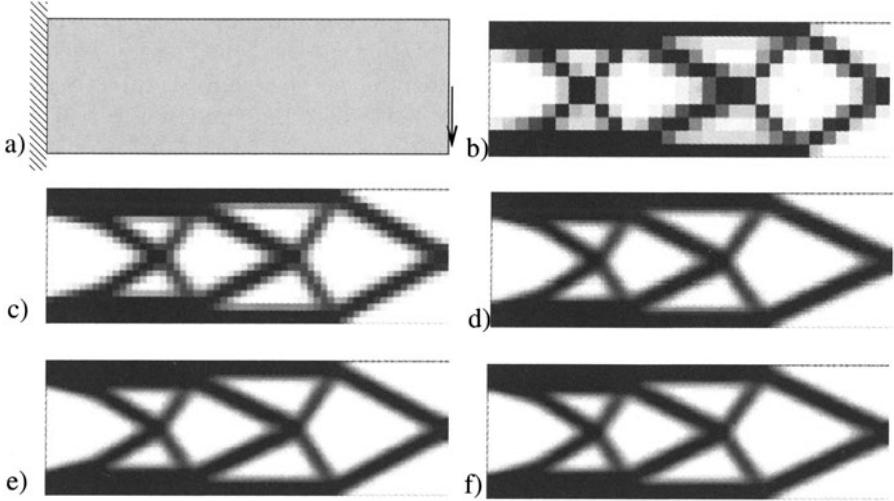


Fig. 1.17. Mesh-independent solutions of the cantilever problem using filtering of sensitivities. a) Design domain and load, b) 300, c) 600, d) 4800, e) 10,800 and f) 19,200 element discretization. Filter radius is 8.2% of the height of the design domain.

The monotonicity of the density can be measured by applying the functional

$$\mathcal{M}(F) = \int_a^b |F'(x)| dx - \left| \int_a^b F'(x) dx \right| \geq 0 , \quad (1.29)$$

which is *zero* if the smooth function F of one variable is monotonic on the interval $[a, b]$, and strictly positive otherwise. As the length scale criterion is violated if the design at any point and at any of the test directions is non-monotonic, one obtains a constraint function $\mathcal{M}_d(\rho)$ with the required properties by “summing” the values of $\mathcal{M}(\rho|_\gamma)$ over all points and all directions γ . We write this as

$$\mathcal{M}_d(\rho) = \int_{\Omega} \left(\sum_{\gamma \in \Lambda(x, d)} \mathcal{M}(\rho|_\gamma) \right)^q d\Omega , \quad (1.30)$$

where $\Lambda(x, d)$ is the set of four diagonals of length d that we test over. The exponent q is used to assure good numerical behaviour of the constraint $\mathcal{M} = 0$ and experiments have shown that $q = 4$ is a good choice in implementations¹³.

In the discrete formulation of the constraint one works with differences of element values of the density ρ and, as for the perimeter constraint, one

¹³ In the actual implementation one uses $\mathcal{M} \leq \delta$, where δ is decreased during the iterative procedure.

has to use a smoothed approximation to the absolute value (cf., (1.21)). We note that the computational effort in evaluating the constraint is linear in the number of elements, and that derivatives can be computed directly (and analytically). The idea also extends to 3-D, where the window is a ball and where one will check along 13 diagonals.

We remark that, as we consider only four directions for checking the minimum length scale, it is clear that there will be some directional dependence of the minimum allowed width of a strip of either material or void. Even in the continuous formulation, the constraint $\mathcal{M} = 0$ only allows piecewise straight lines as boundaries between solid and void. It also restricts the angle between these piecewise straight lines where they meet and the angle between the boundary lines at a kink will be less than $\pi/4$. This shows that more test directions will restrict the boundary curves even further and in the limit of testing along all diagonals only straight lines would be possible as boundaries between solid and void. Finally, we note that, as we cannot make angles sharper than $\frac{\pi}{4}$, the smallest inscribed polygon of a hole is an octagon which for a minimum length d gives the minimum (approximate) radius of curvature $\frac{1+\sqrt{2}}{2}d \approx 1.2d$.

Comparison of methods The perimeter, local gradient and filter methods produce very similar designs, but there are some differences. The perimeter control and the global gradient control schemes are global constraints and will allow the formation of locally very thin bars (albeit in limited numbers). The local gradient and filtering schemes are local constraints and will generally remove thin bars.

Predicting the value of the perimeter constraint for a new design problem must be determined by experiments, since there is no direct relation between local scale in the structure and the perimeter bound. If the perimeter bound is too tight, there may be no solution to the optimization problem. This problem is particularly difficult for three-dimensional problem. In contrast, the gradient and filtering schemes define a local length scale under which structural variation is filtered out. This local length scale corresponds to a lower limit on bar/beam widths. Such a possibility of imposing a minimum length scale is not only of importance for obtaining methods that are stable under mesh-refinement. Almost of greater importance is the possibility this gives for taking manufacturing considerations (machining constraints) into account. This can be in the form of minimum member size requirements for the material phase. This is important for the fabrication of MEMS (Micro-ElectroMechanical Systems, see section 2.6), where mechanisms are etched or deposited by chemical processes. Also, minimum size of a void inclusion is crucial if a structure is machined out by milling processes.

Finally, we remark that the use of a *fixed, finite dimensional set* of designs is a direct way of assuring existence of solutions as well as stability with respect to mesh-refinement – the latter here then *only* means improving the

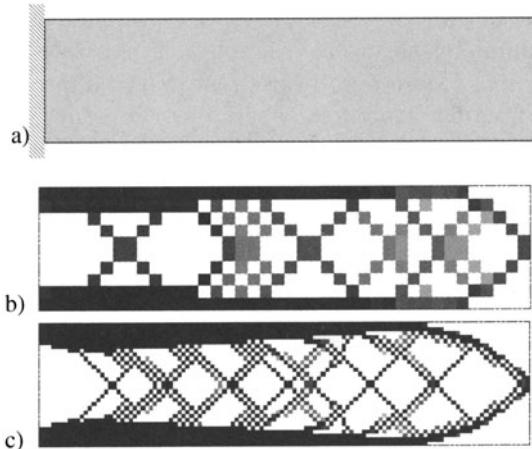


Fig. 1.18. The checkerboard problem demonstrated on a long cantilever beam. a) Design problem, b) solution for 400 element discretization and c) solution for 6400 element discretization.

analysis grid. The geometric resolution cannot be improved beyond what is contained in the initial design description.

1.3.2 The checkerboard problem

Patches of checkerboard patterns appear often in solutions obtained by a direct implementation of the material distribution method that use the displacement based finite element method, cf., figure 1.18. Within a checkerboard patch of the structure the density of the material assigned to contiguous finite elements varies in a periodic fashion similar to a checkerboard consisting of alternating solid and void elements. Such patterns are also observed in the spatial distribution of the pressure in some finite element analyses of Stokes flows. It is now well understood that also for topology design the origin of the checkerboard patterns is related to features of the finite element approximation, and more specifically is due to bad numerical modelling that overestimates the stiffness of checkerboards [9].

The restriction methods already described also has the effect that checkerboarding is reduced or removed . The reason for this is that when one enforces a constraint on geometry (generally speaking in terms of the length of the boundary or in terms of gradient variation) that assure that solutions exist, one also obtains FE-convergence and checkerboards cannot be present for a fine enough mesh (more precisely, they can be made arbitrarily weak).

There are situations where one does *not* wish to enforce a fixed scale geometric restriction on the designs. This is the case when one uses numerical methods to obtain an understanding of the behaviour of optimal topologies at

a fairly fine scale, but in a macroscopic representation. This is of theoretical interest for obtaining insight in for example solutions to problems involving Michell type continua. Moreover, it has great practical interest when designing low volume fraction structures, where one can gain very useful insight for the design of truss and frame structures – here the continuum topology design methodology can predict both member sizes and nodal positions. Another situation where geometric restriction is unwanted is in the computational implementation of the relaxed form of the topology optimization problem, where composites are used to achieve existence of solutions. Also, when studying variable thickness sheets and problems where all possible elasticity tensors are part of the design space, checkerboard control should be achieved by other means than by geometric restriction.

In the following we shall for 2-D problems outline explanations for the appearance of checkerboards and describe a number of methods that can be used to avoid them for cases where geometric scale is not restricted.

The stiffness of checkerboards The most direct explanation as to why checkerboards appear in topology design is that such lay-outs of material have an artificially high stiffness when analyzed in certain discretized formulations. Thus it turns out that a checkerboard of material in a uniform grid of square Q4 elements has a stiffness which is comparable to the stiffness of a $\rho = 1/2$ variable thickness sheet, for any applied loads (or prescribed strains) (see Fig. 1.19 and section 3.3.5). For the minimum compliance problem of an infinite medium, this means that for a Q4 discretization of displacements and any discrete as well as the continuum description of ρ , the corresponding optimization problem has the checkerboard version (matched to the Q4 mesh) as an optimal design. Thus it is not surprising that one in general sees that optimization generates these non-physical checkerboards when Q4-displacement elements are used.

Checkerboards and choice of FE spaces The problem of finding the optimal topology by the material distribution method is a *two* field problem. It involves finding the optimal distribution of material described by the density ρ (or stiffness tensor E) as well as the displacement field u of this optimal design. It is in this connection useful to remember that the displacement based minimum compliance problem we consider can be cast in the form

$$\max_{E \in \mathbb{E}_{\text{ad}}} \min_{v \in U} \left\{ \frac{1}{2} \int_{\Omega} E_{ijkl} \varepsilon_{ij}(v) \varepsilon_{kl}(v) d\Omega - l(v) \right\}, \quad (1.31)$$

and a numerical implementation operates on a discretized version of this min-max type problem for a functional of two variables. It is well-known (cf., Stokes flow), that the finite element analysis of such problems can cause problems, being unstable and being prone to the development of checkerboard patterns for one of the fields. The so-called Babuska-Brezzi (B-B) condition has been developed as a criterion that will guarantee that a

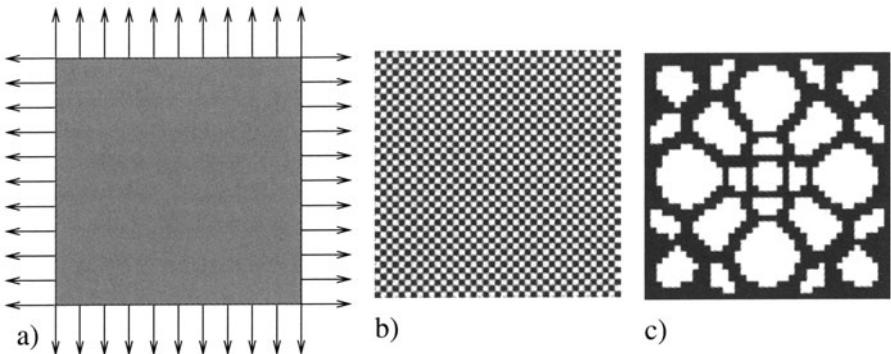


Fig. 1.19. The checkerboard problem demonstrated on a square structure subject to biaxial stress and modelled by Q4 elements. a) Design problem, b) solution without checkerboard control and c) solution with sensitivity filtering. All volume fractions are 50% and the resulting compliances for a variable thickness plate ($p = 1$ in SIMP) (a) is 2.67; for the checkerboard structure (b) 2.81; and for the non-checkerboard structure (c) it is 6.16. Even in this finite lay-out the non-physical checkerboard – modelled by Q4 elements – is almost as stiff as the sheet.

finite element discretization results in a stable numerical scheme, see Brezzi & Fortin (1991). Unfortunately, the functional (1.31) of the topology design problem is not quadratic in the *two* fields and it is also not concave-convex. Thus one cannot directly apply standard saddle point theory and the related application of the Babuska-Brezzi condition to the present situation. However, these problems aside, taking a direct analogy to the similar problem in Stokes flow indicates nonetheless that certain combinations of finite element discretizations will be unstable and some stable. This has been confirmed by numerical experiments for both the SIMP model, for cases with composites and for variable thickness sheets [9]. The analogy suggests that the use of higher order finite elements for the displacement function is a viable method to avoid the checkerboard problem and checkerboards are typically prevented when using 8 or 9-node quadrilaterals for the displacements in combination with an element wise constant discretization of density. An analysis based on a patch test of the finite element models substantiates this finding (Jog & Haber 1996). These patch tests are based on a B-B type analysis of a linearized, incremental form of the necessary conditions, corresponding to an incremental, quadratic approximation of the saddle point problem (1.31), and the tests give information on the performance of various combinations of finite element approximations of the two field problem at hand. We also note that it is possible to extend the full mathematical analyses of mixed FE developed for the Stokes' flow problem to the variable thickness sheet problem (cf., Petersson (1999a)).

The use of higher order finite elements in topology design results in a substantial increase in CPU-time, even though this is not today a serious problem for 2-D problems. But it is still productive to employ alternative and computationally more economical methods. Many such methods have been proposed and typically include some flavour of a mesh related filtering of the densities. A series of such methods will be described below. Before we turn our attention to these concepts, it is worth mentioning some very recent ideas that work with modifications of the typical element density based parametrization.

One is to change the discretization of the density field to be given by the nodal values of the squares that define the mesh for the displacements; the element density is then the *average* of the nodal values (Hammer 2001). A sensitivity of compliance with respect to one of these densities will then depend on the energies in the four neighboring elements, and the design description is in nature similar to filtering methods (see sections 1.3.1). It can be shown that for a finite element discretization based on square elements, this idea corresponds to imposing a local gradient constraint as in 1.23, where G is equal to two times the element size. This means that there always will be a rim of at least one grey element between solid and void elements. Obviously, this also means that this nodal based averaging technique does not imply mesh-independence. An example of the scheme applied to compliant mechanism design is shown in Fig. 2.27. Note that a scheme that uses a density interpolation of nodal values does *not* have the desired effect.

Another idea is to use non-conforming elements for the displacement fields, effectively giving correct zero stiffness to an infinite checkerboard also in the discretized problem (Jang, Jeong, Kim, Sheen, Park & Kim 2003).

Finally, we remark that theoretical studies of the appearance of checkerboards in three-dimensional problems are yet to be carried out. However numerical experience shows that checkerboards also appear for this case.

Removing checkerboards in a patch In order to save CPU-time, but still obtain checkerboard free designs, it has been suggested to employ a patch technique inspired by a method applied for the similar problems in Stokes flow (cf., Johnson & Pitkäranta (1982)). This technique has in practical tests shown an ability to damp the appearance of checkerboards. The strategy controls the formation of checkerboards in meshes of 4-node quadrilateral displacement elements coupled with constant material properties within each element. Thus one maintains the use of low order elements. However, the end result is the introduction of some type of element with a higher number of nodes, as the method in effect results in a “super-element” for the density and displacement functions in 4 neighbouring elements. In what follows we will assume that the design domain Ω is rectangular. It is discretized using a uniform mesh of square, 4-node iso-parametric elements K_{ij} , $i = 1, \dots, 2M$, $j = 1, \dots, 2N$ where $2M$ and $2N$ are the (even) number of elements per side. Consider now, for odd i and j , a patch P_{ij} of four con-

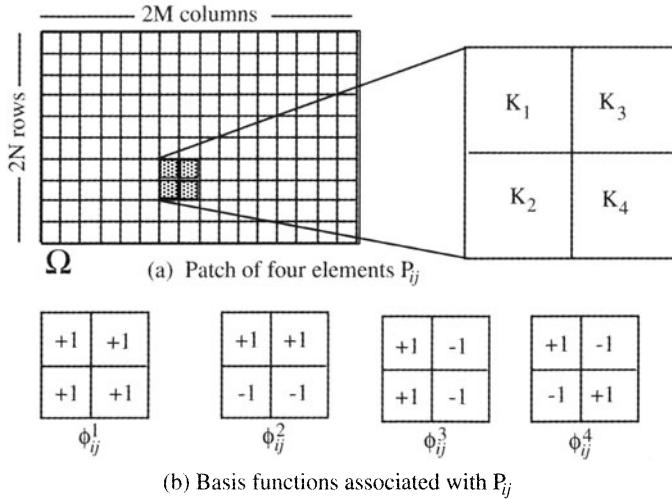


Fig. 1.20. Patches and basis functions used for checkerboard control.

tiguous elements $K_1 = K_{i,j}$, $K_2 = K_{i+1,j}$, $K_3 = K_{i,j+1}$ and $K_4 = K_{i+1,j+1}$, as shown in Fig. 1.20, i.e.,

$$P_{ij} = K_1 \cup K_2 \cup K_3 \cup K_4 .$$

Associated with P_{ij} we introduce basis functions ϕ_{ij}^1 , ϕ_{ij}^2 , ϕ_{ij}^3 and ϕ_{ij}^4 which take the values ± 1 in P_{ij} according to the pattern shown in Fig. 1.20 and are zero outside P_{ij} . Here we note that:

- The functions $\{\phi_{ij}^k\}$ constitute an orthogonal basis,
- A "pure" checkerboard pattern is of the form $u = \sum_{P_{ij}} u_{ij} \phi_{ij}^4$.

This suggests that in order to avoid the formation of checkerboard patterns we need to restrict ρ to lie within the more restricted, checkerboard-free space

$$\bar{V} = \left\{ v \mid v(x) = \sum_{P_{ij}} (v_{ij}^1 \phi_{ij}^1 + v_{ij}^2 \phi_{ij}^2 + v_{ij}^3 \phi_{ij}^3), (v_{ij}^1, v_{ij}^2, v_{ij}^3) \in \mathbf{R}^3, \right. \\ \left. i = 1, 3, \dots, 2N - 1, j = 1, 3, \dots, 2M - 1 \right\} .$$

This restriction on ρ links the four elements in a patch, and the amount of material in $K_1 \cup K_4$ equals that of $K_2 \cup K_3$ and each is half of the total volume of the patch.

The coupling of the density distribution makes it difficult to apply the usual iterative optimality condition method. In MMA one can work directly with the design space \bar{V} by using the $3MN$ parameters $(v_{ij}^1, v_{ij}^2, v_{ij}^3)$ as design variables. This, however, changes the simple bound constraints $0 \leq \rho \leq 1$ into a huge number of linear constraints on the parameters v , making this option impractical. Instead, the following simpler procedure, which has been applied in a variety of problems, can be employed for both algorithms:

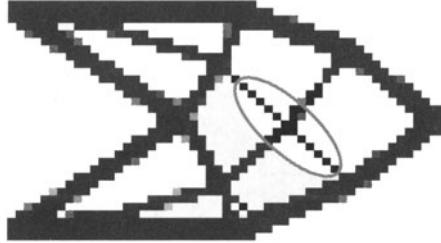


Fig. 1.21. Single checkerboard-patterns may form within the patch control scheme. But no extended checkerboard patterns can be present (from Poulsen 2002b).

1. At each iteration of the optimization algorithm the cell size parameters within each element K_{ij} are updated using the usual update method (optimality criteria approach or an MMA step).
2. For each patch P_{ij} let $\{\rho_1, \rho_2, \rho_3, \rho_4\}$ be the updated densities in the four quadrants of the patch associated with the updated cell sizes (using the numbering of 1.20). We then seek, as the starting point for the next iteration, a new piece-wise constant and *checkerboard-free* density distribution within the patch, say $\bar{\rho}$, written as

$$\bar{\rho}(x) = \frac{1}{4}(\rho_1 + \rho_2 + \rho_3 + \rho_4)\phi^1 + \bar{v}_2\phi^2 + \bar{v}_3\phi^3, \quad x \in P_{ij}.$$

Here $\bar{\rho}$ is *checkerboard-free* (as $\bar{v}_4 = 0$) and it preserves material in the patch (as the coefficient of ϕ^1 is set as $\bar{v}_1 = \frac{1}{4}(\rho_1 + \rho_2 + \rho_3 + \rho_4)$). To determine the parameters \bar{v}_2, \bar{v}_3 , we select $\bar{\rho}$ as the best L^2 approximation to ρ in P_{ij} under the constraints that $0 \leq \bar{\rho}_i \leq 1, i = 1, 2, 3, 4$. This corresponds to a QP problem in two variables, with linear constraints. The solution can be found analytically, and is given as

$$\begin{aligned}\bar{\rho}_1 &= \frac{1}{4}(3\rho_1 + \rho_2 + \rho_3 - \rho_4), & \bar{\rho}_3 &= \frac{1}{4}(\rho_1 - \rho_2 + 3\rho_3 + \rho_4), \\ \bar{\rho}_2 &= \frac{1}{4}(\rho_1 + 3\rho_2 - \rho_3 + \rho_4), & \bar{\rho}_4 &= \frac{1}{4}(-\rho_1 + \rho_2 + \rho_3 + 3\rho_4),\end{aligned}$$

if these values satisfies $0 \leq \bar{\rho}_i \leq 1$. If a $\bar{\rho}_i$ in these expressions is above 1, it is set to 1 and the corresponding diagonal density is adjusted to maintain the volume of the patch; negative values are handled likewise and are set to 0. The modification of the density outlined here has the flavour of a filtering in a post-processing step that is invoked at each step of the optimization procedure and should therefore be used with some caution. We note that it does not disturb areas of the domain where no checkerboard control is needed, and also remark again that the method corresponds to introducing a "super-element" of four Q4 elements with a total of 9 displacements nodal points and with 3 degrees of freedom for the density approximation. Thus

the method maintains more resolution in densities, as compared to, say, the approach of using Q9 elements for displacements and element-wise constant density ρ .

An alternative to the procedure above is to perform a change of variables that allows one to work directly with checkerboard free designs. Inspired by work in wavelet-based parametrization of the design (see section 1.5.5), we introduce a checkerboard free space of auxiliary variables $\bar{W} = \{w \in \bar{V}\}$. We do not impose any side constraints on $w(x)$ and convert it into to a density satisfying the bounds $0 < \rho < 1$ by a transformation:

$$\rho(x) = h(w(x)), \text{ with } h(w) = \frac{\arctan(w)}{\pi} + \frac{1}{2},$$

where h is a strictly increasing function¹⁴ that will map checkerboard free patches of the auxiliary variable w to checkerboard free patches of the density ρ .

In the optimization, the variables $(w_{ij}^1, w_{ij}^2, w_{ij}^3)$ then become the design variables. For implementation, sensitivity information with respect to these variables are needed, but this information can be gained from sensitivities wrt. $\rho(x)$ by an application of the chain-rule. The volume constraint becomes a non-linear function in the auxiliary variables, but this does not create any difficulties. Finally one notes that box constraints on the auxiliary variables are a requirement of MMA (as in most mathematical programming algorithms) and these can be chosen big enough as not to affect the results; also, one can use an imposed w_{\min} to match a desired value of ρ_{\min} , but this is not critical [11].

We note here that the schemes proposed above depend on the build-up of the mesh in 2 by 2 patches of quadrilaterals. In each such patch checkerboards are removed, but checkerboards between patches are still possible if the row or column number of the upper left corner of a checkerboard is even. This also means that corner to corner patterns of the single elements can occur, but no large areas of checkerboards are possible. This is discussed in further detail in section 2.6 that deals with design of mechanisms.

NoHinge: A checkerboard constraint In section 1.3.1 geometry control was achieved by defining one extra constraint for the optimization problem. This idea can also be implemented for checkerboard control, i.e., one defines a non-negative constraint function that should have value *zero* for the design to be free of checkerboards.

Consider the patch of square elements in figure 1.22. Defining the function

$$m(x, y, z) = |y - x| + |z - y| - |z - x|,$$

that is zero if the sequence of real numbers x, y, z is monotonic (increasing, decreasing or constant) and strictly positive otherwise, we can determine

¹⁴ A function $S(w) = \frac{1}{2} \frac{1-e^{-kw}}{1+e^{-kw}} + \frac{1}{2}$ is also a choice among many other possibilities.

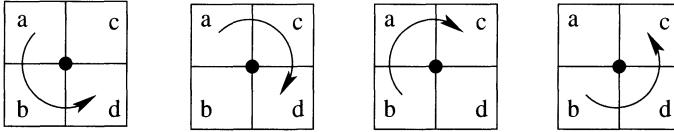


Fig. 1.22. A check for monotonicity along four paths around an interior node.

that the patch is free of checkerboard patterns, if just one of the numbers $m(a, b, d)$, $m(a, c, d)$, $m(b, a, c)$ or $m(b, d, c)$ is zero. This can be in turn be expressed as the condition that the number

$$h(a, b, c, d) = m(a, b, d) m(a, c, d) m(b, a, c) m(b, d, c) ,$$

is zero. A design defined by a density ρ that is element wise constant on a mesh of quadrilaterals with N interior nodes will thus be free of checkerboards if it satisfies the constraint

$$\sum_{k=1}^N h(\rho_{k,a}, \rho_{k,b}, \rho_{k,c}, \rho_{k,d}) = 0 , \quad (1.32)$$

where $\rho_{k,e}$, $e = a, b, c, d$ are the material densities in the elements connected to the node k . This constraint can thus be added to our optimization problem to assure checkerboard free solutions. It can also be used to remove “artificial” hinges in mechanism design, see section 2.6. As we have seen in other situations, an implementation using gradient based optimization techniques requires a replacement of the absolute value by a smooth substitute, for example $|x| \simeq \sqrt{x^2 + \epsilon^2} - \epsilon$ with $\epsilon = 0.1$. With this modification a sensitivity analysis of the constraint is straightforward, but rather tedious (Poulsen 2002a). For an example of the use of this scheme, see Fig. 2.25.

Checkerboard control by filtering of sensitivities The filtering technique for gradients described in 1.3.1 can also be cast in a version that only constrains checkerboards, without imposing a mesh independent length scale. This just requires that one adjusts the filter in (1.27) to exactly making the design sensitivity of a specific element depend on a weighted average over the element itself and its *eight* direct neighbours. This is a is very efficient method for removing checkerboards [9].

1.3.3 Non-uniqueness, local minima and dependence on data

It is important to observe that most problems in topology design (as in structural problems in the large) are not convex. Moreover, many problems have multiple optima, i.e. *non-unique solutions*. An example of the latter is the design of a structure in uni-axial tension. Here a structure consisting of one thick bar will be just as good as a structure made up of several thin bars with

the same overall area. The non-convexity typically means that one can find several different local minima (which is what the gradient based algorithms locate) and one can obtain different solutions to the same discretized problem when choosing different starting solutions and different parameters of the algorithms. Most global optimization methods seem to be unable to handle problems of the size of a typical topology optimization problem. Based on experience, it seems that continuation methods must be applied to ensure some sort of stable convergence towards reliably good designs.

The idea of *continuation methods* is to gradually change the optimization problem from an (artificial) convex (or quasiconvex) problem to the original (non-convex) design problem in a number of steps. In each step a gradient-based optimization algorithm is used until convergence. This is useful in many types of problems. Examples are the use of a continuation method where the structure first is optimized allowing regions consisting of composites (see Chap. 3), and after convergence, a penalization scheme is gradually introduced to obtain a 0-1 design. Likewise (as mentioned in section 1.2.4), for SIMP it is advisable to start out with $p = 1$ and then slowly raise the value of p through the computations until the final design is arrived at. For the perimeter constraint it is also beneficial to perform a gradual decrease of value of the constraint on the perimeter. For the mesh-independence filter (see section 1.3.1) it is normally recommended to start with a large value of the filter size r_{min} (which gives designs with blurry edges) and to gradually decrease it, to end up with a well-defined 0-1 design.

Finally, it is extremely important to observe that the results that one obtains with topology design of course depends on the data that one decides on using before applying the optimization procedure. Thus a change of the geometry of the design domain, the choice of load and boundary conditions can result in drastical changes in the “optimal design” that an algorithm may produce. Similar effects can be seen from variations of perimeter constraint values or filter parameters, etc. This is actually not that surprising as we are dealing with very “nasty” optimization problems, but in topology design this effect is just much more noticeable than in many other types of structural optimization problems.

1.4 Combining topology and shape design

Traditionally, in shape design of mechanical bodies, a shape is defined by the oriented boundary curves or boundary surfaces of the body and in shape optimization the optimal form of these boundaries is computed. This approach is very well established and the literature is extensive [2], [35]. On the other hand, we have just seen how the material distribution formulation can give a good estimate of the boundary of a structure, but here a reasonable prediction of the finer details of the boundaries requires very large FEM models. Also, the inherent large scale nature of the topology optimization method is

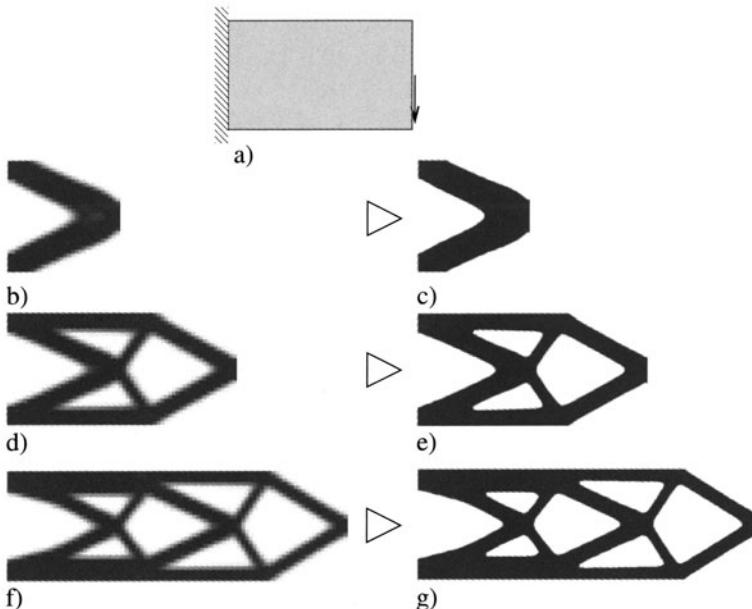


Fig. 1.23. Postprocessing of grey-scale pictures by automatic (MATLAB) contourplotting. Cantilever beam for different aspect ratios. b), d) and f): optimized topologies based on SIMP and filtering of sensitivities and c), e) and g): contour plots based on the grey-scale pictures.

such that the objectives used for the optimization should be global criteria, e.g. compliance, volume, average stress, etc., so that the effectiveness of the dual optimizers can be maintained by treating problems with a moderate number of constraints. For example, the focal point in the presentation so far has been the minimization of the compliance of a structure subject to a constraint on the volume of the structure. On the other hand, the description of the body by boundary curves and surfaces allows the finer details of the body to be controlled by a moderate number of design variables (e.g., spline control points) so this setting is better suited for studying problems such as the minimization of the maximum value of the displacements or of the Von Mises equivalent stress in the body¹⁵.

It is thus for this type of situations natural to integrate the material distribution method and the boundary variations approach into one design tool, employing the topology optimization techniques as a pre-processor for boundary shape optimization. The possibility of generating the optimal topology for a body can be used by the designer to select the shape of the initial proposed form of the body for the boundary variations technique. This is usually

¹⁵ The handling of local stress constraints for continuum topology design problems is described in section 2.3.

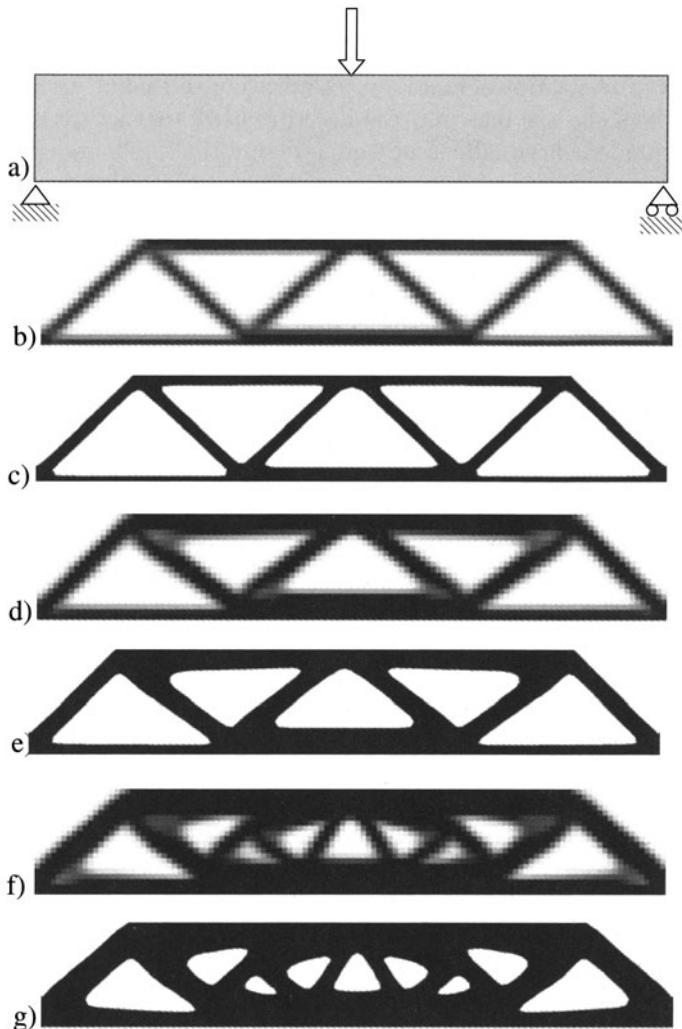


Fig. 1.24. Postprocessing of grey-scale pictures by automatic (MATLAB) contourplotting. MBB-beam for volume fractions of 30%, 50% and 70%, respectively. b), d) and f): optimized topologies based on SIMP and filtering of sensitivities and c), e) and g): contour plots based on the grey-scale pictures. Only moderate modifications (smoothing of corners etc.) seems to be needed before manufacturing.

left entirely to the designer, but the material distribution method gives the designer a rational basis for his choice of initial form. As to be expected, the topology is of great importance for the performance of the structure, and it has turned out that - not unexpectedly - the compliance optimized topologies generated using topology design are very good starting points for optimiza-

tion concerning several other criteria such as maximum stress, maximum deflection, etc.

The direct integration of topology optimization and shape design methods is made difficult by the fact that the description of a structure by a density function is fundamentally different from a description by boundary curves or surfaces, as used in boundary variations shape optimization methods. In a CAD integrated shape optimization system, it is perhaps natural that the integration is based on the designer drawing the initial shape for the boundary variations technique directly on the top of a picture of the topology optimized structure, allowing for designer interaction [10]. This also creates a design situation where the ingenuity of the designer is put to use for generating a “good” initial form from the topology optimization results. The term “good” in this context covers considerations such as ease of production, aesthetics, etc. that may not have a quantified form. However, automatic interfacing between the topology optimization method and other structural optimization methods is no doubt more productive. Here image processing and smooth surface generation are key technologies [10]. Such techniques are especially important for an effective integration of topology design methodology in general purpose 3-D Computer Aided Optimal Design (CAOD) systems. We note that any integration of the two design methods is simplified by the fact that the integration can be based on a common FEM mesh generator and analysis module and a common CAD input-output facility. The requirements on the mesh generator are mainly governed by the boundary variations technique, as mesh distortions and mesh non-uniformities for that problem can become critical due to the shape changes of the analysis domain.

An important aspect of shape design is adaptivity of the FE mesh. Likewise, for topology design one can work with a sequence of design situations where the groundstructure (the reference domain) as well as the FE mesh is subject to adaptation; for further details consult [10].

It should be emphasized that the boundary variation method in essence is computationally significantly more involved than the topology design method. Also, the mathematical technicalities of formulating the problem and computing sensitivity information are more daunting, as is indicated in appendix 5.3. On the other hand, the material distribution method is a large scale optimization problem. Describing boundaries by for example spline control points requires a much lower number of design variables, meaning that standard mathematical programming techniques can be used also for problems with a substantial number of constraints. Note also that the basic approach to topology design is of equal complexity for two and three dimensional structures, but that the description of geometry for boundary shape design is much more complicated in dimension three.

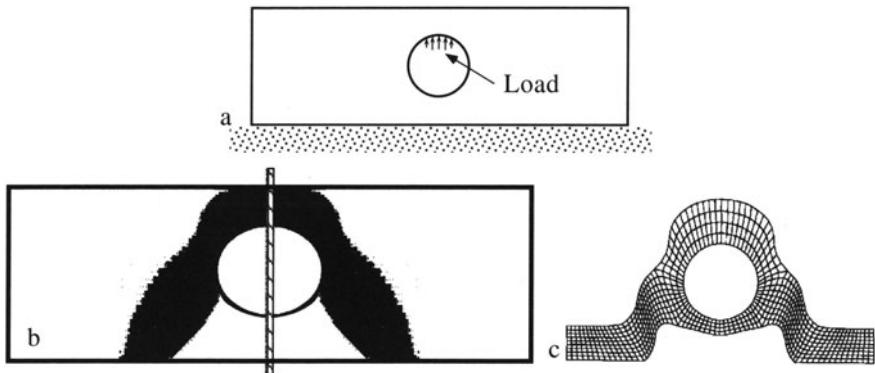


Fig. 1.25. The optimal design of a bearing pedestal, using the homogenization approach (see Chap. 3) integrated with the boundary shape design system CAOS (see Rasmussen et al. (1993)). a) The reference domain, with loading. The rim of the inner hole was kept as a solid in the topology optimization. b) The result of the homogenization approach. c) The final design, after boundary shape design for minimum maximal Von Mises stress and after adding outer parts to the structure for fastening. Utilizing symmetry only one half of the structure was analysed, as indicated in b) (from Olhoff et al. 1992a).

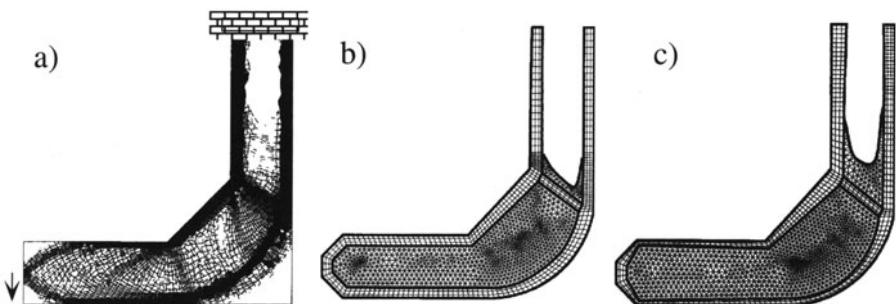


Fig. 1.26. Optimized topology and shape design of a structure made of two materials, resulting in a sandwich structure. a) Optimized two-material topology computed using rank-3 layered materials (see Chap. 3). b) Initial design for a refinement using boundary shape optimization. All boundaries between skin and core are restricted to be piecewise straight lines. For the boundary design the weight is minimized without increasing the compliance relative to the optimal topology. c) Final shape optimized structure. By courtesy of Rasmussen, Thomsen and Olhoff.

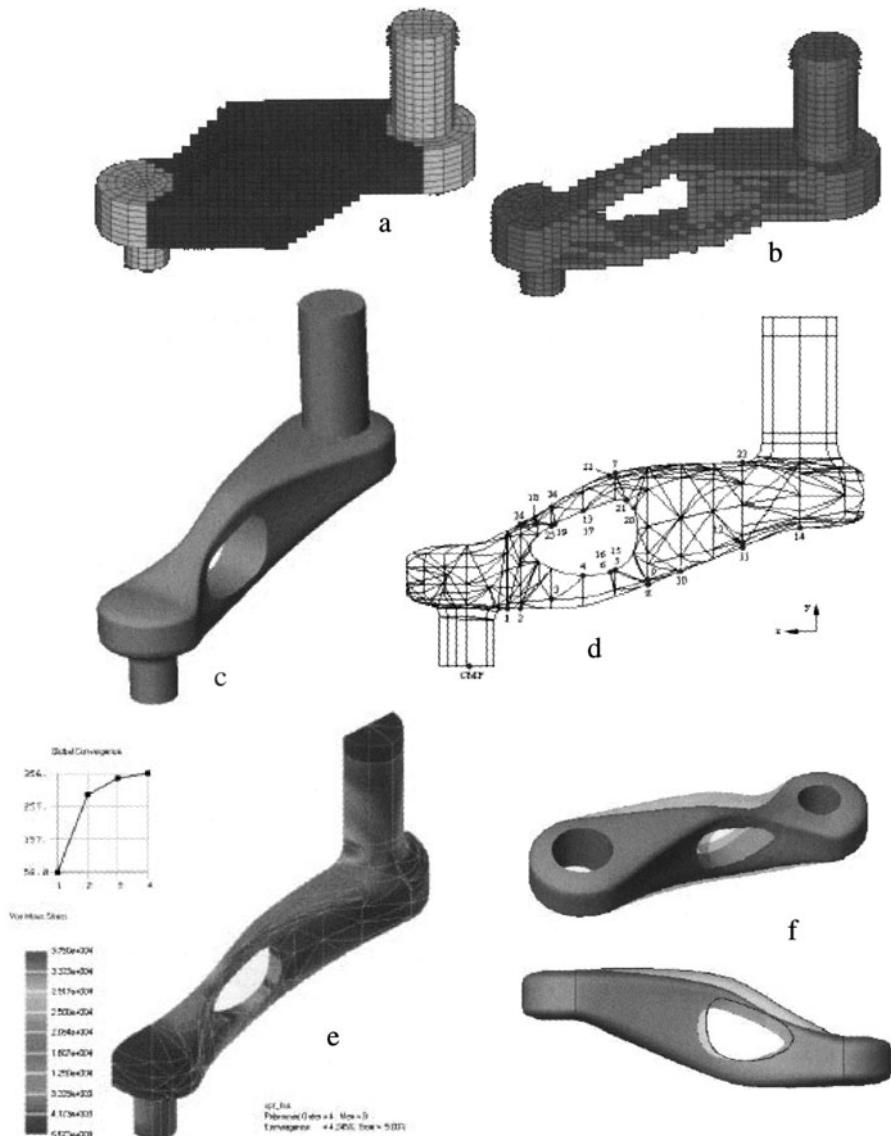


Fig. 1.27. Integrated optimal design of a vehicle roadarm. a) Initial Finite Element Model, b) topology optimized roadarm, c) reconstructed solid model, d) Finite Element mesh for shape design e) Von Mises stress of the shape optimized design and f) comparison of the 3D Roadarm before and after shape design (Light grey: initial design, dark grey: optimized design) (from Tang & Chang 2001).

1.5 Variations of the theme

1.5.1 Multiple loads

The framework described for minimum compliance design for a single load case generalizes easily to the situation where design for multiple load conditions is formulated as a minimization of a weighted average of the compliances for each of the load cases. We here obtain a simple multiple load formulation as:

$$\begin{aligned} \min_{u^k \in U, E} & \sum_{k=1}^M w^k l^k(u^k) \\ \text{s.t. : } & a_E(u^k, v) = l^k(v), \quad \text{for all } v \in U, k = 1, \dots, M, \\ & E \in \mathbb{E}_{\text{ad}}, \end{aligned} \tag{1.33}$$

for a set w^k, f^k, t^k, l^k , $k = 1, \dots, M$, of weighting factors, loads and tractions, and corresponding load linear forms given as

$$l^k(u) = \int_{\Omega} f^k u \, d\Omega + \int_{\Gamma_T^k} t^k u \, ds,$$

for the M load cases we consider.

In this formulation the displacement fields for each individual load case are independent, thus implying that the multiple load formulation for the displacement based case has the equivalent form

$$\begin{aligned} \max_{E \in \mathbb{E}_{\text{ad}}} & \min_{\hat{u} = \{u^1, \dots, u^M\}} \left\{ \int_{\Omega} W(E, \hat{u}) \, d\Omega - l(\hat{u}) \right\}, \\ W(E, \hat{u} = \{u^1, \dots, u^M\}) &= \frac{1}{2} \sum_{k=1}^M w^k E_{ijpq}(x) \varepsilon_{ij}(u^k) \varepsilon_{pq}(u^k), \\ l(\hat{u} = \{u^1, \dots, u^M\}) &= \sum_{k=1}^M w^k l^k(u^k). \end{aligned} \tag{1.34}$$

Likewise, we have a stress based formulation

$$\min_{E \in \mathbb{E}_{\text{ad}}} \min_{\substack{\operatorname{div} \sigma^k + f^k = 0 \text{ in } \Omega, \\ \sigma^k \cdot n = t^k \text{ on } \Gamma_T^k \\ k=1, \dots, M}} \left\{ \frac{1}{2} \int_{\Omega} \sum_{k=1}^M w^k C_{ijpq} \sigma_{ij}^k \sigma_{pq}^k \, d\Omega \right\}. \tag{1.35}$$

For the stiffness modelled as in the SIMP model, the optimality criteria method developed for the single load case generalizes directly and we obtain an update scheme for ρ_K at iteration step K which is exactly the same as given in (1.12) of section 1.2.1, but with a modified “energy” expression

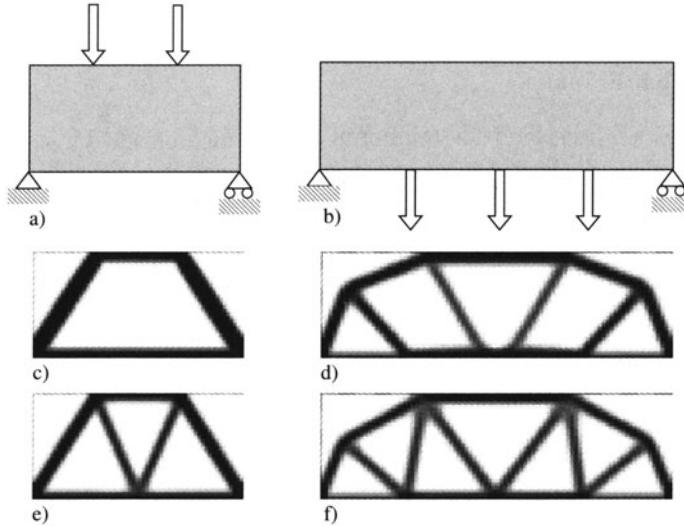


Fig. 1.28. Example of differences in using one or more load cases. a) and b) Design domains. c) and d) Optimized topologies for all loads in one load case. e) and f) Optimized topologies for multiple loading cases. It is seen that single load problems result in unstable structures based on square frames whereas multi load case problems results in stable structures based on triangular frames.

$$B_K = \Lambda_K^{-1} p \rho(x)^{(p-1)} E_{ijnm}^0 \sum_{k=1}^M w^k \varepsilon_{ij}(u_K^k) \varepsilon_{nm}(u_K^k) .$$

Similarly, for use of an algorithm like MMA, the sensitivity of the weighted average of compliances just becomes the weighted average of the sensitivities of each of the compliances. Also, the similarity of the iterations in MMA and in the optimality criteria method remains. Finally, it may be remarked that the inclusion of extra load cases is very cheap since the stiffness matrix already has been factorized.

1.5.2 Variable thickness sheets

For planar problems, the stiffness tensors given by the SIMP method reduces to the setting of the well-known variable thickness sheet design problem if we set $p = 1$; in this circumstance the density function ρ is precisely the thickness h of the sheet. The minimum compliance problem then becomes

$$\begin{aligned} \min_{u,h} \quad & l(u) \\ \text{s.t. : } & a_h(u, v) \equiv \int_{\Omega} h(x) E_{ijkl}^0 \varepsilon_{ij}(u) \varepsilon_{kl}(v) d\Omega = l(v), \text{ for all } v \in U, \quad (1.36) \\ & \int_{\Omega} h(x) d\Omega \leq V, \quad h_{\min} \leq h \leq h_{\max} < \infty. \end{aligned}$$

Problem (1.36) can also be written as (cf., (1.6))

$$\begin{aligned} \min_{\substack{h \in L^\infty(\Omega), \\ h_{\min} \leq h \leq h_{\max} < \infty \\ \int_{\Omega} h(x) d\Omega \leq V}} \quad & c(h) \\ c(h) = \max_{v \in U} \quad & \left\{ 2l(v) - \int_{\Omega} h(x) E_{ijkl}^0 \varepsilon_{ij}(v) \varepsilon_{kl}(v) d\Omega \right\}. \end{aligned} \quad (1.37)$$

As the stiffness is linear in h , the compliance c is convex, as it is given as a maximization of convex functions. Also, the complete problem statement (1.37) is a convex-concave saddle point problem that (as noted earlier) lends itself to a complete FE convergence analysis (see Petersson (1999b)) within the framework of the theory developed for the Stokes' flow problem. The variable thickness sheet design problem also corresponds very closely to truss design problems in the sense that the stiffness of the structure as well as the volume of the structure depend linearly on the design variable for both models. This implies that a discrete version of the problem can be solved using some very efficient algorithms that have been developed for truss topology design (cf. Chap. 4). These algorithms do not require that $h_{\min} > 0$, and the setting thus allows for a prediction of the the optimal topology of the sheet without the ambiguity inherent in the chosen value of h_{\min} ; this is especially important here as we do not force the design towards a 0-1 design.

The linear dependence of the stiffness on the design function h has an even more significant implication for the continuum problem, as one can prove existence of solutions (see appendix 5.2.1). Thus there is no need for restriction methods or the introduction of materials with micro structure (this holds for minimization of compliance and optimization of the fundamental frequency). Finally, we remark that the variable thickness sheet problem also plays a significant role when considering optimal design within a completely free parametrization of the stiffness tensors over all positive definite tensors in 2-D as well as 3-D. Here the problem form (1.36) arises after a reduction of the original full formulation; this will be discussed in detail in Chap. 3.

Explicit penalization of thickness The variable thickness design problem has been used as the inspiration for topology design methods where one seeks the optimum over all isotropic materials with given Poisson ratio and linearly varying Young's modulus [11]. This formulation results in designs with large domains of "grey" and modifications are necessary to obtain 0-1 designs. This can be accomplished by adding to the objective an explicit penalty of

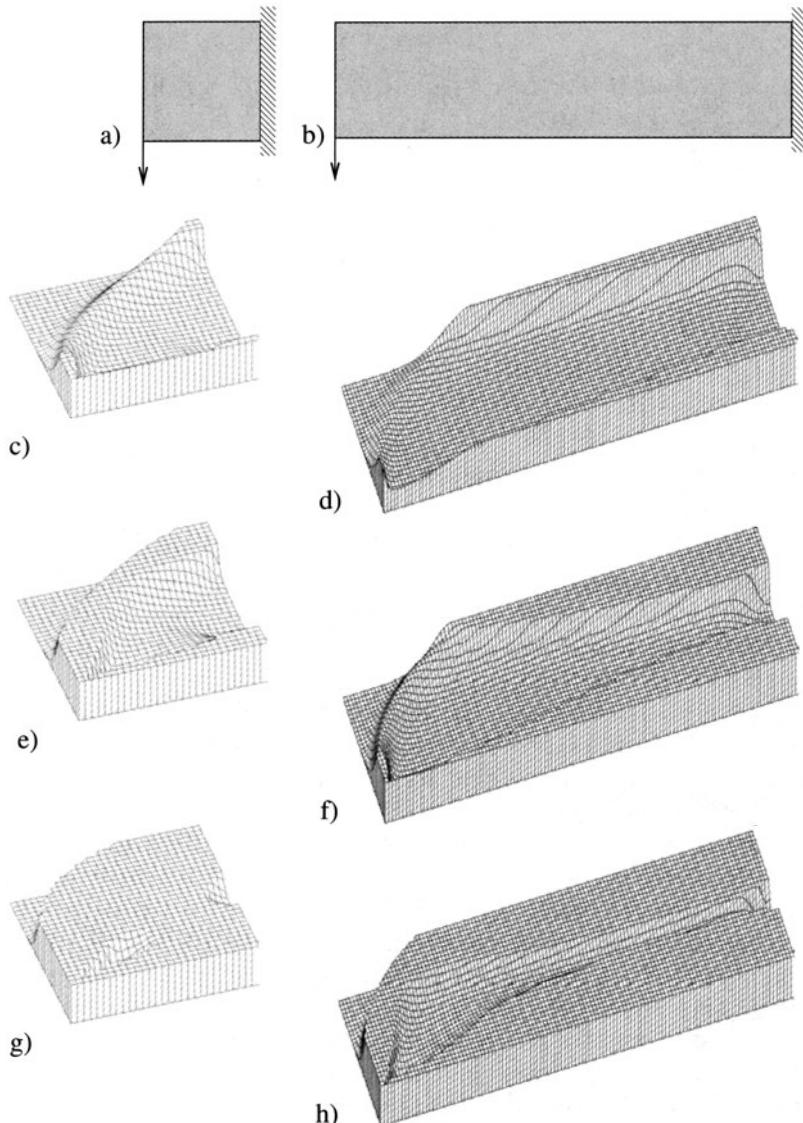


Fig. 1.29. The design of a variable thickness sheet for two cantilever-like ground structures with aspect ratios a) 1:1 and b) 1:4. c) – h): The optimal designs for a volume constraint c) and d) 30%, e) and f) 60% and g) and h) 90%, respectively, of the volume of a design with uniform thickness h_{\max} (cf. constraints on thickness). Notice that the areas of intermediate thickness is considerable, especially for low amounts of available material. Thus the variable thickness design does not predict the topology of the structure as a true 2-dimensional object, but utilizes that the structure is in effect a 3-dimensional object.

intermediate densities, for example in the form of functionals (we revert to using a density ρ as the design variable):

$$\mathcal{W}(\rho) = \int_{\Omega} \rho(x)(1 - \rho(x))d\Omega, \quad \hat{\mathcal{W}}(\rho) = \int_{\Omega} \rho(x)^2(1 - \rho(x))^2d\Omega. \quad (1.38)$$

Alternatively, the penalty function can be used as a constraint $\mathcal{W}(\rho) \leq \delta$ for some small $\delta \geq 0$.

The use of a penalty function such as \mathcal{W} (or $\hat{\mathcal{W}}$) has a detrimental effect on the very nice mathematical properties of the original variable thickness sheet problem. For one, existence of solutions is no longer true. However, existence of solutions can be recovered (Borrvall & Petersson 2001b) by modifying the penalty function (1.38) to the form

$$\widetilde{\mathcal{W}}(\rho) = \int_{\Omega} (\rho * K)(x)(1 - (\rho * K)(x))d\Omega, \quad (1.39)$$

where one evaluates the original penalty function on a *filtered* version of the density ρ (we use here the notation introduced in section 1.3.1). The filter smoothes the density before penalization and as such provides for a more severe penalization than does \mathcal{W} (for details, consult Borrvall & Petersson (2001b)). Thus using $\widetilde{\mathcal{W}}$, the designs become almost entirely black and white (a 0-1 design) if the penalty factor is large enough (see Fig. 1.12).

The penalty approach just outlined maintains the existence of solutions for the problem of minimum compliance and the maximization of the fundamental frequency. If a broader range of problems is to be considered the restriction techniques, as described in section 1.3.1, should be applied. Note also that the use of the penalty $\widetilde{\mathcal{W}}$ makes it impossible to use the efficient truss-type algorithm mentioned above (but an MMA inspired optimality criterion method is an efficient alternative (Borrvall & Petersson 2001a)).

In order to maintain the structure of the original computational problem it has been suggested instead to consider a sequence of problems where the volume constraint in each step K of the sequence is modified as (see for example Guedes & Taylor (1997), Rodrigues, Soto & Taylor (1999))

$$\int_{\Omega} w_K(x)\rho(x)d\Omega \leq V,$$

where the weight function w_K is fixed and determined from the optimal solution ρ_{K-1} to the prior step so as to penalize low density regions:

$$w_K(x) = \begin{cases} T_K & \text{if } \rho_{K-1} \leq \delta \\ 0 & \text{otherwise.} \end{cases}$$

With suitable big values of T_k and a small value of δ this scheme generates 0-1 designs and each step is computationally equivalent to the original variable thickness sheet problem. In implementation, the tuning of the penalization becomes an issue. Also note that as above, the advantages of this idea is closely linked to the properties of the minimum compliance problem.

1.5.3 Plate design

Studies of the problem of variable thickness plate design and the appearance of stiffeners in such design problems have played a crucial role in the developments in optimal structural design [29]. Thus topology design and especially design with materials with microstructure can be seen as a natural extension of the original work by Cheng and Olhoff on plates. In this sense this exposition of topology design methods is reversed relative to history, but it is today more natural to consider plate design as a special variation of the general framework.

The design of variable thickness Kirchhoff plates or Mindlin plates is at first glance just another sizing problem of finding the optimal continuously varying thickness of the plate. The close connection with the 0-1 topology design problems is not entirely evident, but the cubic dependence of plate bending stiffness on the thickness of the plate implies that the optimal design prefers to achieve either of the bounds on the thickness, in essence a plate with integral stiffeners. This in turn implies non-existence of solutions unless the gradient of the thickness function is constrained or the problem is extended to include fields of infinitely many stiffeners; this latter concept is dealt with in Chap. 3.

Variable thickness design of Kirchhoff plates The minimum potential energy statement for a Kirchhoff plate is of the form¹⁶

$$\min_w \left\{ \frac{1}{2} \int_{\Omega} \frac{h^3}{12} E_{ijkl}^0 \kappa_{ij}(w) \kappa_{kl}(w) d\Omega - \int_{\Omega} f w d\Omega \right\},$$

where f is the transverse load. The thickness of the plate is denoted by h and we assume that the mid-plane is a plane of symmetry. The deformation of the plate is described by the transverse displacement of the mid-plane w , with associated (linearized) curvature tensor $\kappa_{ij} = \frac{\partial^2 w}{\partial x_i \partial x_j}$, and the relationship between the curvature tensor and moment tensor M is given as

$$M_{ij} = D_{ijkl} \kappa_{kl}, \quad \text{with} \quad D_{ijkl} = \frac{h^3}{12} E_{ijkl}^0,$$

where E_{ijkl}^0 is the plane stress elasticity tensor of the given material. The similarity between the curvature-moment relation for plates and the strain-stress relation in elasticity hides the fundamental difference that the Kirchhoff plate is governed by a fourth order scalar equation, while standard linear elasticity is governed by a system of second order equations. As for the variable thickness sheet problem, the thickness of the plate also here automatically provides the plate design problem with a continuous design variable. Considering the minimization of compliance the most natural problem to consider is thus

¹⁶ Here and elsewhere in this section (section 1.5.3) all indices range over 1 and 2.

$$\max_{D \in \mathbf{PE}_{ad}} \min_w \left\{ \frac{1}{2} \int_{\Omega} D_{ijkl} \kappa_{ij}(w) \kappa_{kl}(w) d\Omega - \int_{\Omega} f w d\Omega \right\}, \quad (1.40)$$

with the set \mathbf{PE}_{ad} of bending stiffnesses given as

$$\begin{aligned} D_{ijkl} &= \frac{h^3}{12} E_{ijkl}^0, \quad h \in L^\infty(\Omega), \\ 0 \leq h_{\min} &\leq h \leq h_{\max} < \infty, \quad \int_{\Omega} h d\Omega \leq V, \end{aligned} \quad (1.41)$$

which looks like a SIMP interpolation scheme! Also in the plate setting, problem (1.40) with the design set (1.41) is not well posed, and the existence of solution is not always assured. This was first vividly demonstrated by Cheng & Olhoff (1981), who discovered the formation of stiffeners in numerically computed “optimal” solutions for high ratios of h_{\max}/h_{\min} and h_{\max}/h_{unif} , where $h_{\text{unif}} = V / \int_{\Omega} d\Omega$, see Fig. 3.21 in Chap. 3. The number of stiffeners increase when the discretization of design is refined, with a resulting (substantial) decrease in compliance, a situation completely similar to the behaviour of the 0-1 topology design setting. Compared to the variable thickness design problem for sheets, this is caused by the cubic dependence of the stiffness of the plate on the thickness. Physically, this dependence makes it advantageous to move as much material as possible away from the mid-plane of the plate, for example in the form of integral stiffeners. A method to obtain mesh-independence and existence of solutions is analogous to what has been described in section 1.3.1, by restricting the variation of the thickness function, for example in the form of a constraint on the slope (gradient) of the thickness function. Example solutions with a point wise bound on the slope of the thickness of a rotational symmetric plate were first shown in Niordson (1983).

The computational procedure for computing optimal plate designs is completely analogous to the procedure described earlier in this chapter and the optimality criteria and sensitivity calculations carry over ad verbatim, with strains and stresses interpreted as curvatures and moments, respectively.

Topology design for Mindlin plates We close this brief discussion on plate design by considering some models for the design of Mindlin plates.

The minimum potential energy statement for a *constant thickness* Mindlin plate constructed from one material is of the form

$$\min_u \left\{ \frac{1}{2} \int_{\Omega} h E_{ijkl}^0 \varepsilon_{ij}(u) \varepsilon_{kl}(u) d\Omega + \frac{1}{2} \int_{\Omega} \frac{h^3}{12} E_{ijkl}^0 \kappa_{ij}(u) \kappa_{kl}(u) d\Omega \right. \\ \left. + \frac{1}{2} \int_{\Omega} h D_{ij}^S \gamma_i(u) \gamma_j(u) d\Omega - \left(\int_{\Omega} f u d\Omega + \int_{\Gamma_i} t u d\Gamma \right) \right\},$$

where f is the transverse and t the in-plane load. The thickness of the plate is denoted by h and we assume that the mid-plane is a plane of symmetry. Also, E_{ijkl}^0 is the plane stress elasticity tensor and D^S is the transverse shear stiffness matrix. In Mindlin plate theory generalized displacements of the plate $u = (u_1, u_2, w, \theta_1, \theta_2)$ consist of the in-plane displacements (u_1, u_2) , the fibre

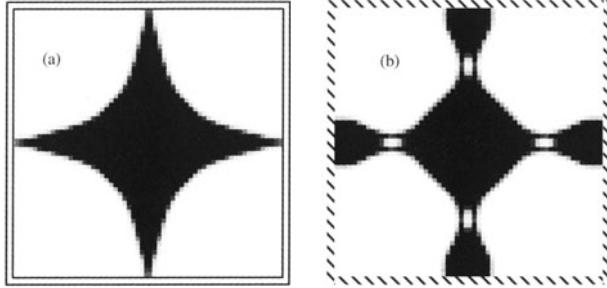


Fig. 1.30. Resulting topologies for compliance minimization of square Mindlin plates. The material volumes are restricted to 25% of the filled plates and the plates are loaded with a force at the center. a) Simply supported and b) clamped plate (from Pedersen 2001).

rotations (θ_1, θ_2) and the transverse displacement of the mid-plane w . The associated membrane, bending, and transverse shear strains are, respectively,

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad \kappa_{ij} = \frac{1}{2} \left(\frac{\partial \theta_i}{\partial x_j} + \frac{\partial \theta_j}{\partial x_i} \right), \quad \text{and} \quad \gamma_i = \frac{\partial w}{\partial x_i} - \theta_i .$$

For plates there are several options for performing topology design, connected to the possibility to also consider out-of-plane variations of the build-up of the plate. For the design of a perforated plate one would thus use thickness functions that attain values 0 or \bar{h} , for example implemented with the help of a density function ρ and a SIMP interpolation:

$$h = \rho^p \bar{h}, \quad \text{Vol} = \int_{\Omega} \rho \bar{h} d\Omega .$$

Other possibilities is to consider *reinforcement* of a given plate or to consider the design of a sandwich structure, where two outer skins are given and the topology design deals with the topology design of the inner core [29].

1.5.4 Other interpolation schemes with isotropic materials

The use of SIMP or the penalized variable thickness formulation have in the last few years been supplemented by some alternative interpolation schemes that have certain theoretical or computationally advantageous features for specific problems. As they fall within the class of interpolation models with isotropic materials we briefly discuss them in this chapter. The use of composites is the theme of Chap. 3.

Hashin-Shtrikman bounds The so-called Hashin-Shtrikman bounds for two-phase materials express the limits of *isotropic* material properties that

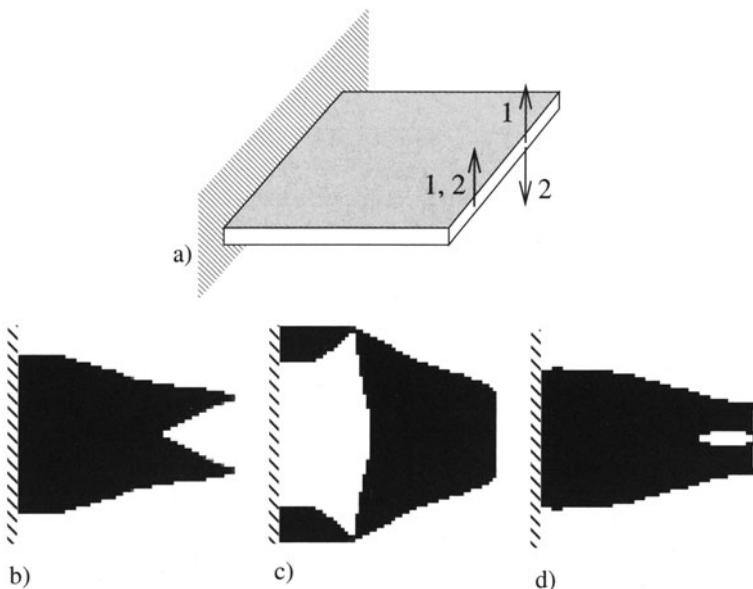


Fig. 1.31. Resulting topologies for compliance minimization of square Mindlin cantilever plates. a) Design problem with loads for the first and second load case. b) Solution to one-load case problem with two upward oriented forces, c) solution to one-load case problem with one force upwards and one downwards and d) solution to two load case problem. The volume fraction is 50% (by Niels L. Pedersen).

one can possibly achieve by constructing composites (materials with microstructure) from two (or more) given, linearly elastic, isotropic materials [4]. These bounds give expressions of material parameters as functions of volume fraction, or for our purposes as functions of density ρ of material, and can thus be employed as interpolation schemes (all material laws involved will be isotropic). For our purposes we work with two materials, one with a low stiffness E^{\min} and one with high stiffness E^0 . The corresponding values of the Poisson ratios are ν^{\min} and ν^0 .

The Hashin-Shtrikman bounds are typically expressed in terms of the bulk and shear moduli of the materials, κ and μ (corresponding to the eigenvalues of the stiffness tensor). Restricting ourselves here to 2-D plane elasticity, we have for isotropic materials that

$$\kappa = \frac{E}{2(1-\nu)}; \quad \mu = \frac{E}{2(1+\nu)} \quad (\text{in } 2-D).$$

The bounds are then in terms of these parameters (in 2-D)¹⁷ (we assume here that $\kappa^0 \geq \kappa^{\min}$ and $\mu^0 \geq \mu^{\min}$):

$$\kappa_{\text{upper}}^{HS} = (1 - \rho)\kappa^{\min} + \rho\kappa^0 - \frac{(1 - \rho)\rho(\kappa^{\min} - \kappa^0)^2}{(1 - \rho)\kappa^0 + \rho\kappa^{\min} + \mu^0}, \quad (1.42)$$

$$\mu_{\text{upper}}^{HS} = (1 - \rho)\mu^{\min} + \rho\mu^0 - \frac{(1 - \rho)\rho(\mu^{\min} - \mu^0)^2}{(1 - \rho)\mu^0 + \rho\mu^{\min} + \frac{\kappa^0\mu^0}{\kappa^0 + 2\mu^0}}, \quad (1.43)$$

$$\kappa_{\text{lower}}^{HS} = (1 - \rho)\kappa^{\min} + \rho\kappa^0 - \frac{(1 - \rho)\rho(\kappa^{\min} - \kappa^0)^2}{(1 - \rho)\kappa^0 + \rho\kappa^{\min} + \mu^{\min}}, \quad (1.44)$$

$$\mu_{\text{lower}}^{HS} = (1 - \rho)\mu^{\min} + \rho\mu^0 - \frac{(1 - \rho)\rho(\mu^{\min} - \mu^0)^2}{(1 - \rho)\mu^0 + \rho\mu^{\min} + \frac{\kappa^{\min}\mu^{\min}}{\kappa^{\min} + 2\mu^{\min}}}. \quad (1.45)$$

Each combination of formulas $\kappa_{\text{upper}}^{HS}$, μ_{upper}^{HS} and $\kappa_{\text{lower}}^{HS}$, μ_{lower}^{HS} represents an interpolation of the material properties of the two materials, and any convex combination is also an interpolation scheme, which then satisfies the bounds. Thus a whole range of schemes can be generated. Here the *lower* bound interpolation penalizes intermediate densities most. Note that the Hashin-Shtrikman bounds represent materials that have both a Young's modulus and a Poisson ratio that vary with density (even if the two base materials have the same Poisson ratio).

If both materials have Poisson ratio equal to 1/3, then the upper and lower bounds (and all convex combinations) also represent a material law with Poisson ratio $\nu = 1/3$, and the bound can be expressed in terms of the Young's modulus only:

$$\text{for } \nu = 1/3 : \begin{cases} E_{\text{upper}}^{HS} = E^0 \frac{\rho E^0 + (3 - \rho)E^{\min}}{(3 - 2\rho)E^0 + 2\rho E^{\min}}, \\ E_{\text{lower}}^{HS} = E^{\min} \frac{(2 + \rho)E^0 + (1 - \rho)E^{\min}}{2(1 - \rho)E^0 + (1 + 2\rho)E^{\min}}. \end{cases} \quad (1.46)$$

This reduces further if the weak material is void ($E^{\min} = 0$):

$$E_{\text{upper}} = \frac{\rho E^0}{3 - 2\rho}, \quad E_{\text{lower}} = \begin{cases} 0 & \text{for } \rho < 1, \\ E^0 & \text{for } \rho = 1, \end{cases} \quad (1.47)$$

which is then an interpolation with void and with a material with $\nu = 1/3$ and Young's modulus E^0 . For many test cases in topology design one works within this framework of $\nu = 1/3$.

¹⁷ The bounds are *necessary conditions*. It is known that not all combinations of numbers κ and μ that satisfy the bounds actually represent the bulk and shear moduli of a realizable material - see Chap. 2.

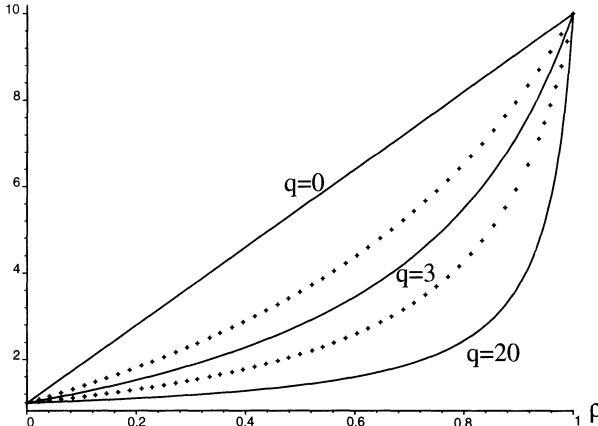


Fig. 1.32. The interpolation using RAMP (solid curves) compared with the Hashin-Shtrikman bounds (dotted curves).

SIMP and the Hashin-Shtrikman bound If we require that an interpolation model in any sense can be related to a composite made of the given materials, then we should demand that the model satisfies the Hashin-Shtrikman bounds stated above. For SIMP one of the material phases is zero, i.e., $E^{\min} = 0$. Then the only relevant Hashin-Shtrikman bound (1.42) simplify somewhat and it is possible to show that SIMP satisfies the bounds if the power of the model satisfies the inequalities stated in (1.5) (Bendsøe & Sigmund 1999). As already noted, this does not assure that a composite can actually be constructed. But we shall in Chap. 2 see how topology design (sic!) can be used to construct microstructures that realizes the SIMP interpolation scheme.

An approach with rational functions In the section on the variable thickness sheet problem it was seen how the expression of the compliance c via the potential energy makes it possible to conclude that the compliance in that situation is convex. On the other hand, if c is derived in terms of the complementary energy (cf., (1.7), we have that

$$c = \min_{\sigma \in S} \left\{ \int_{\Omega} C_{ijkl} \sigma_{ij} \sigma_{kl} d\Omega \right\}. \quad (1.48)$$

This shows that if one can make the compliance tensor C_{ijkl} a concave function of the design, then c , as a minimization of concave functions, becomes concave. And this is advantageous if we want 0-1 designs that are on the “border” of the space of densities (this will be made precise below). For interpolating the compliance we cannot work with vanishing stiffness, but instead operate with a low stiffness E^{\min} and interpolate between this and the properties E^0 (we assume constant Poisson ratio). The simplest concave

interpolation of the inverse of the Young's modulus E is then the linear one (as proposed in Stolpe & Svanberg (2001b) and Rietz (2001)), which is

$$\frac{1}{E(\rho)} = \frac{1}{E^{\min}} + \rho \left(\frac{1}{E^0} - \frac{1}{E^{\min}} \right).$$

This, in turn, corresponds to the following rational expression for E :

$$E(\rho) = E^{\min} + \frac{\rho}{1 + q(1 - \rho)} (E^0 - E^{\min}), \quad (1.49)$$

where $q = \bar{q}$, $\bar{q} \equiv (E^0 - E^{\min})/E^{\min}$ (the corresponding expression for SIMP is $E = E^{\min} + \rho^p (E^0 - E^{\min})$). The relation given in (1.49) can also be used for other values of q ; the scheme has in certain circumstances been given the acronym RAMP for Rational Approximation of Material Properties.

The interpolation (1.49) makes, by construction, the compliance a concave function of ρ if q is chosen as $q = \bar{q}$ (or bigger). Moreover, $q = 0$ gives the linear interpolation (which makes compliance convex), and for materials with Poisson ratio 1/3 the interpolation equals the Hashin-Shtrikman *upper* bound for $q = 2(E^0 - E^{\min})/(E^0 - 2E^{\min})$ and the Hashin-Shtrikman *lower* bound for $q = \frac{2}{3}\bar{q}$. The intermediate densities are thus severely penalized for larger values of q , cf., figure 1.5.4. A natural continuation method using this interpolation scheme is then to begin the optimization procedure with $q = 0$ and then increase q until $q \geq \bar{q}$. The concavity of compliance for such large values of q implies the existence of a globally optimal 0-1 solution for a FE discretized version of the problem, for example where we use element wise constant densities in a mesh of squares, only the simple bounds $0 \leq \rho \leq 1$ and a volume constraint that is an integer times the volume of the base element (Stolpe & Svanberg 2001b). This has the added side effect that one can conclude similarly for the SIMP interpolation¹⁸ provided we choose $p \geq (\bar{q} + 1)$. If for example a perimeter constraint is added to the problem statement this property does not hold any longer; however, the intrinsic penalization of the interpolation still results in designs that are almost free of grey, if q is large enough. It is also worth noting that if E^{\min} is much smaller than E^0 (as typical for finding the topology of a structure made from one material), then the “magic” value \bar{q} becomes large (and infinite for the limit of $E^{\min} = 0$).

We remark here that if design of two-material structures is the goal of the topology optimization, the RAMP model is in a sense more physical than SIMP. The latter will always violate the Hashin-Shtrikman bounds for small density values, while RAMP has a whole range of q values for which the bound is satisfied (for a Poisson ratio 1/3). However, RAMP does not satisfy these bounds for the range of q where the compliance becomes concave.

¹⁸ This follows from the property that for a given ρ , the SIMP interpolation gives a higher compliance than the RAMP interpolation, when $p \geq q + 1$.

A Reuss-Voigt interpolation scheme The Voigt upper bound for the effective properties of a mixture of two materials state that for any strain field ε we have that the strain energy W of the composite is bounded from above by the expression

$$W \leq [\rho E_{ijkl}^0 + (1 - \rho) E_{ijkl}^{\min}] \varepsilon_{ij} \varepsilon_{kl},$$

where the two materials have elasticity tensors E^0 and E^{\min} , respectively, and where the volume fraction of the material E^0 is ρ . Likewise, the Reuss lower bound states that the energy W is bounded from below by the expression

$$W \geq [\rho C_{ijkl}^0 + (1 - \rho) C_{ijkl}^{\min}]^{-1} \varepsilon_{ij} \varepsilon_{kl},$$

where C denotes the compliance tensors of the materials.

These two bounds can be combined to a convex combination to what has been named a Reuss-Voigt interpolation scheme (Swan & Arora 1997, Swan & Kosaka 1997a)

$$E_{ijkl}^{VR}(\rho) = \alpha [\rho E_{ijkl}^0 + (1 - \rho) E_{ijkl}^{\min}] + (1 - \alpha) [\rho C_{ijkl}^0 + (1 - \rho) C_{ijkl}^{\min}]^{-1}. \quad (1.50)$$

Here α is a parameter which weighs the contributions from the Voigt and Reuss bounds. If one of the materials is void, the interpolation introduces a jump (discontinuity) at $\rho = 1$ which is not present if both materials have some stiffness. In that case, for two materials that both have a Poisson's ratio of $\nu = 1/3$ we have that the Hashin-Shtrikman bounds are satisfied if and only if $\alpha = 1/3$ (Bendsøe & Sigmund 1999).

Spline-based approach The SIMP interpolation scheme has zero slope at zero density. Thus the stiffness converges to zero orders of magnitude faster than mass and this has proven to be a difficulty when considering vibration problems. For benign computational behaviour in these problems (see section 2.1) one needs an interpolation scheme where the ratio of mass to stiffness ($\rho/E(\rho)$) remains finite in the limit of vanishing density ρ . The RAMP scheme and the Hashin-Shtrikman bound schemes have this feature. Another way to secure this property is to construct a Bézier curve interpolation in the (ρ, E) -plane that connects the two points $(0, 0)$ and $(1, E^0)$ and has tangents $(1, k_1)$ and $(1, k_2)$, respectively, at the endpoints. As shown in Pedersen (2002c), such a Bézier curve with four control points can be given a parametrization as

$$\left. \begin{aligned} \rho(t) &= \frac{1 - k_1}{k_1 - k_2} (3t - 3t^2) + t^3 \\ E(t) &= k_1 \frac{1 - k_1}{k_1 - k_2} (3t - 3t^2) + t^3 \end{aligned} \right\}, \quad t \in [0, 1]. \quad (1.51)$$

For a given density, the first equation gives the corresponding parameter value t , and this in turns gives the stiffness E . For (1.51) to satisfy the Hashin-Shtrikman bounds for (Poisson ratio 1/3) one needs to choose the slopes so that $k_1 \leq 1/3$ and $k_2 \geq 3$, and as for the other schemes, almost 0-1 designs are best achieved through a continuation method. Here that means working with decreasing values of k_1 and increasing values of k_2 .

1.5.5 Design parametrization with wavelets

The possibility of controlling geometric features of the designs that are obtained from topology design has already been a central theme in section 1.3.1. There such control is basically achieved by filtering techniques or by imposing constraints on the pixel (voxel) based description of design. Another possibility should be to work with alternative design descriptions that inherently allows for some form of control of geometric complexity. This means that one will express the density function ρ in terms of basis functions where the coefficients will govern aspects of the geometry. A natural choice for such a representation is *wavelets*, as these can represent data that is localized in space as well as in frequency [11].

The general framework of wavelet representation of image data will not be treated here. Instead we will indicate how one from a pixel representation can construct an alternative design description that directly works with data at different scales¹⁹. We take as the starting point the setting of the patch-based checkerboard control described in section 1.3.2. As a first step of changing the design representation one chooses to describe the density distribution in the domain as

$$\rho(x) = \sum_{P_{ij}} (v_{ij}^1 \phi_{ij}^1 + v_{ij}^2 \phi_{ij}^2 + v_{ij}^3 \phi_{ij}^3 + v_{ij}^4 \phi_{ij}^4),$$

where the basis functions ϕ_{ij}^k , $k = 1, 2, 3, 4$, are illustrated in Fig. 1.20. In this expansion the coefficients v_{ij}^1 represents the average value of the densities in the patch, and the other coefficients express the local variations from this average. One can now consider the larger scale mesh consisting of the patches P_{ij} and the corresponding average values v_{ij}^1 , which can be likened to standing back from a picture so that local variations are averaged out. Here we can again repeat this procedure of expanding in terms of basis functions $\tilde{\phi}_{ij}^k$, $k = 1, 2, 3, 4$, at this larger scale of working with 4-patches of the patches P_{ij} (and the values v_{ij}^1). This procedure is possible to perform if the number of elements per side is 2^M by 2^N , and it can be continued until the last patch consists of the full domain divided into four “super-elements”. In terms of a design representation, we now have parameters that at different levels of

¹⁹ This is based on the discrete Haahr-wavelet and the associated Mallat decomposition.

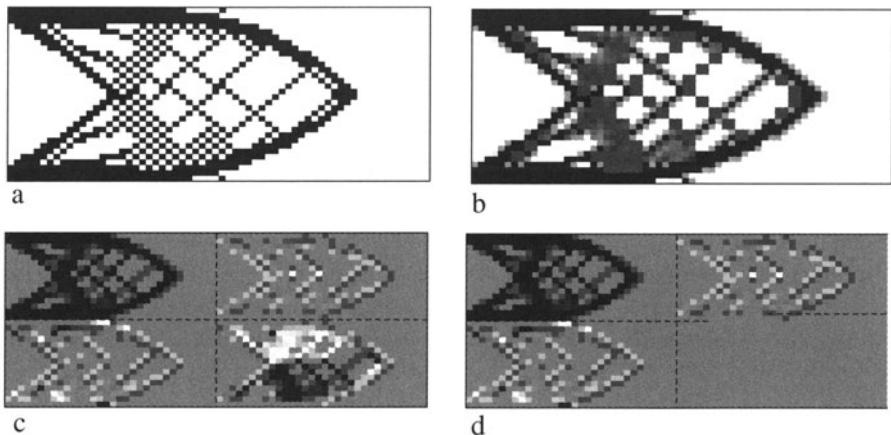


Fig. 1.33. A design with checkerboards, here shown in terms of its a) pixel values and c) in terms of the basis functions ϕ_{ij}^k , $k = 1, 2, 3, 4$: the top left quarter of c) is the design in a spatial resolution which is half the original one, and the three other quarters represent the difference from this to the original design in various ways. The original design in a) can be obtained by combining the four parts. In b), checkerboards have been removed by setting the lower right corner of c) equal to zero, as shown in d) (from Poulsen 2002b).

fineness describe the density in terms of averages and variations from the averages.

The alternative parametrization of design just outlined gives a direct way to control the overall length scale from the level of fineness of the basis functions that are applied. Moreover, it makes it possible to systematically do the design process from coarse scale to fine scale (Kim & Yoon 2000). Finally, it makes it possible to work directly with a checkerboard-free design space, as seen in section 1.3.2. However, the wavelet representation should not be used directly in topology optimization, as it introduces a huge number of constraints associated with the condition $0 \leq \rho \leq 1$ which regrettfully reduces computational efficiency. Instead, it is recommended to use intermediate variables as described in section 1.3.2 (Kim & Yoon (2000), Poulsen (2002b)).

We close this brief sketch by noting that wavelet based methods have also been used as an alternative to finite elements for the analysis part of topology design, see DeRose Jr. & Díaz (1999), DeRose Jr. & Díaz (2000). Moreover, it is expected that use of more advanced wavelet bases than the Haahr-wavelet should potentially lead to more refined methodologies to control geometry.

1.5.6 Alternative approaches

The technique for topology design of continuum structures that is described in this monograph is based on the concept of optimal distribution of material, using interpolation of material properties together with mathematical programming. As we shall see in Chap. 2, this is a universally efficient approach for a broad range of problems in engineering design. In parallel with the development of this methodology, other schemes have also evolved [13]. Some of these work within the same modelling framework using algorithms for *discrete* optimization or various types of growth/shrinking procedures, but a completely different modelling paradigm can also be found in for example the *bubble method*. We will here only briefly mention some of these concepts, and refer to Eschenauer & Olhoff (2001) for a survey.

Solving the discrete problem The introduction of the interpolation schemes for the 0-1 design problem is extremely useful as it allows for the use of mathematical programming methods for continuous (smooth) problems. However, it would be very useful if one could attack the original formulation directly²⁰ [13]. This has been done for the compliance design problem using dual methods, that have been shown to be effective in the absence of local constraints. Methods like simulated annealing or genetic algorithms have also been tested for more general settings, but their need for many function evaluations is computationally prohibitive, but for rather small scale examples (each call involves a costly finite element analysis on a grid at least as fine as the raster representation of the design).

It has been shown recently (Stolpe & Svanberg 2001a) that for a broad class of problems one can formulate the 0-1 topology design problem as a *linear* mixed continuous-integer programming problem and this will no doubt be useful for generating more efficient methods for treating the discrete format in the future.

Growing and shrinking a structure; Bone remodelling Numerous methods have been proposed for dealing with topology design without the use of mathematical programming [13]. They are typically named as “evolutionary” methods, but they are *not* in any way connected to the use of genetic algorithms. On the contrary, these methods typically work with concepts that are similar to the idea of *fully stressed design*, i.e., material is added to highly stressed areas of a design and removed from understressed areas of the design, typically implemented by an addition or removal of elements from the FE model.

Some implementations of such concepts are very similar to an optimality criteria type algorithm, but the removal and adding of elements can lead to erroneous results. This is basically because gradient information is used to

²⁰ In a well-posed form, for example with a perimeter constraint.

perform changes of variables between zero and one, as illustrated in Zhou & Rozvany (2001) for the case of minimum compliance design.

It is interesting to note that many models used in bio-mechanics for bone adaptation have a form which is similar to the optimality criteria algorithm described in section 1.2.1 [7]. These models are usually based on energy arguments and are not derived from an optimization principle. This similarity in approach to material redistribution updates has also lead to bone adaptation models being proposed as topology redesign methods.

Topological variations and level sets The concepts of *topological derivatives* and the *bubble method* is based on utilizing ideas from the boundary variations technique for shape design as a basis for topology design [13].

The topological derivative of a functional as compliance expresses the sensitivity with respect to the opening of a small (infinitesimal) hole at a certain position in the analysis domain. Likewise, in the bubble method a criterion is developed that allows for the prediction of the most effective location for creating a hole and this information is used to perform a boundary variations shape optimization of the resulting topology. The hole placement is then repeated in this shape optimized structure leading to good designs with smooth boundaries. A direct application of the topological derivative in a mathematical programming technique is presently not possible, as there is no evident underlying parametrization available; implementations have thus been based on techniques reminiscent of element removal techniques.

The application of level-set techniques for topology design have also been proposed recently [13]. The contours of a parametrized family of level-set functions are here used to generate the boundaries of a structure, and the topology can change with changes in the level-set function. This technique is in an initial stage of development.